JAM multi-step strategy

Carlota Andrés
Jefferson Lab
A.I. for Nuclear Physics Workshop

In collaboration with: J.J. Ethier, W. Melnitchouk, N. Sato
Traditionally different types of collinear distributions (PDFs, FFs) are obtained from independent analyses.

Performing *simultaneous* fits of different collinear distributions allows us to:

- Study the limits in $x$ and $Q^2$ of collinear factorization
- Test the universality of PDFs, FFs...
- Extract the distributions in a *rigorous* way where all the data are studied using the same theoretical framework

In this talk: (first) *simultaneous* analysis of *unpolarized* PDFs and FFs → *Strange* quark distribution
## Evolution of JAM

<table>
<thead>
<tr>
<th>Process</th>
<th>JAM15</th>
<th>JAM16</th>
<th>JAM17</th>
<th>JAM19</th>
<th>JAM20?</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIS</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DY</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIA</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIDIS</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>JAM15</th>
<th>JAM16</th>
<th>JAM17</th>
<th>JAM19</th>
<th>JAM20?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$\Delta f$</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>$D_h^f$</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$D_f^f$</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

+ More processes
### Evolution of JAM

<table>
<thead>
<tr>
<th>Process</th>
<th>JAM15</th>
<th>JAM16</th>
<th>JAM17</th>
<th>JAM19</th>
<th>JAM20?</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIS</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DY</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIA</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIDIS</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>JAM15</th>
<th>JAM16</th>
<th>JAM17</th>
<th>JAM19</th>
<th>JAM20?</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f )</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \Delta f )</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>( D^h_f )</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
</tr>
</tbody>
</table>

**First simultaneous analysis of unpolarized PDFs and FFs**

+ More processes
## Evolution of JAM

<table>
<thead>
<tr>
<th>Process</th>
<th>JAM15</th>
<th>JAM16</th>
<th>JAM17</th>
<th>JAM19</th>
<th>JAM20?</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIS</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DY</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIA</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIDIS</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>JAM15</th>
<th>JAM16</th>
<th>JAM17</th>
<th>JAM19</th>
<th>JAM20?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$\Delta f$</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>$D^h_f$</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

First **simultaneous** analysis of **unpolarized** PDFs and FFs

**Why JAM19?**

To study the **strange** quark distribution

+ More processes
Motivation II

- The strange PDF is **less known** than the non-strange light flavors.

- Traditionally: **neutrino-(heavy) nucleus** DIS data used to extract the strange PDF.
  - **Drawbacks:** nuclear effects on PDFs.

- **$W$ and $Z$** inclusive production in $p-p$ collisions also sensitive to flavor separation.
  - **Drawbacks:** tension between CMS and ATLAS results?
Motivation II


\[ \mu^2 = 1.9 \text{ GeV}^2, n_f = 3 \]

\[ r_s = \frac{(s + \bar{s})/2}{d} \]

\[ x \]

- CMS
- NuTeV/CCFR + NOMAD + CHORUS
- CHORUS + CMS + ATLAS

AI Workshop
First challenge: a wide variety of data

DIS: $l + (p, d) \rightarrow l' + X$

DY: $l + (p, d) \rightarrow l\bar{l} + X$

SIDIS: $l + d \rightarrow l' + h + X$

SIA: $e^+ + e^- \rightarrow h + X$

Total: 4366 data points!

$Q^2 > m_c^2$

$W^2 > 10 \text{ GeV}^2$
Second challenge: MC fits

• Typical PDF parametrization:

\[ x \Delta f(x) = N x^a (1 - x)^b (1 + c \sqrt{x} + dx) \]

\[ \chi^2 = \sum_{e}^{N_{\text{exp}}} \sum_{i}^{N_{\text{data}}} \frac{(D_{i}^e - T_{i})^2}{(\sigma_{i}^e)^2} \]

• Perform single $\chi^2$-fit: Multiple local minima!
  Parameters difficult to constrain
  Hessian method for uncertainties Introduces tolerance criteria
  Unsuitable for simultaneous analysis of collinear distributions

• Monte Carlo methods:
  • Allow efficient exploration of the parameter space
  • Uncertainties directly obtained from MC replicas
JAM19
multi-step methodology
JAM19: multi-step fitting

PDFs

$x f(x)$

$x$

+ DIS data
JAM19: multi-step fitting

PDFs

$xf(x)$

$x$

+ DIS data

+ DIS + DY data
JAM19: multi-step fitting

PDFs

$x f(x)$

+ DIS data
+ DIS + DY data

PION FF

$z D(z)$

+ SIA pion data

Carlota Andrés
JAM19: multi-step fitting

PDFs

$xf(x)$

$zD(z)$

$x$

+ DIS data
+ DIS + DY data

PION FF

$zD(z)$

$z$

+ SIA pion data

KAON FF

$zD(z)$

$z$

+ SIA kaon data

Carlota Andrés
JAM19: multi-step fitting

PDFs

$x f(x)$

+ DIS data
+ DIS + DY data
+ SIDIS data

$D(z)$

PION FF

+ SIA pion data
+ SIDIS pion data

KAON FF

+ SIA kaon data
+ SIDIS kaon data

Carlota Andrés
Discriminating multiple solutions

\[ f(x) \]

\[ x \]
Discriminating multiple solutions

\[ f(x) \]

\[ x \]
Discriminating multiple solutions

\[ f(x) \]

\[ x \]

\[ x_g \]
Discriminating multiple solutions

\[ f(x) \]

\[ x \]
Discriminating multiple solutions

\[ R_s = \frac{s + \bar{s}}{\bar{u} + d} \]

\( x_f(x) \)

\( x \)
k-means clustering

E.g. \( f(x) = x^\alpha (1 - x)^\beta \)

\((\alpha^*, \beta^*) : \text{centroid}\)

\((\alpha_i, \beta_i) : \text{replica}\)
k-means clustering

E.g. \( f(x) = x^\alpha (1 - x)^\beta \)

\((\alpha^*, \beta^*) : \text{centroid}\)

\((\alpha_i, \beta_i) : \text{replica}\)
k-means clustering

E.g. \( f(x) = x^\alpha (1 - x)^\beta \)

\((\alpha^*, \beta^*) : \text{centroid}\)

Initialization  Assignment  Update

\((\alpha_i, \beta_i) : \text{replica}\)
k-means clustering

E.g. \( f(x) = x^\alpha (1 - x)^\beta \)

\((\alpha^*, \beta^*) : \text{centroid}\)

\((\alpha_i, \beta_i) : \text{replica}\)
k-means clustering

E.g. \( f(x) = x^\alpha (1 - x)^\beta \)

\((\alpha^*, \beta^*) : \) centroid

Initialization

Assignment

Update

Assignment

Repeat until convergence

Carlota Andrés

AI Workshop
Discriminating multiple solutions

+ DIS data
+ DIS + DY data
+ SIDIS data
Discriminating multiple solutions

+ DIS data

+ DIS + DY data

+ SIDIS data
SIA $K^+/K^-$ data

Carlota Andrés
SIA K+/K- data

Data/Theory

$z$

$z_{D_{K^+}}$
SIA $K^+$/K- data

$\bar{s} \rightarrow K^+$
SIDIS K-

Unfavored solutions
Large \( s(x) \)
Small \( D_{s^\pm}^K(z) \)

Favored solutions
Large \( D_{s^\pm}^K(z) \)
Small \( s(x) \)

\[
\chi^2_{\text{SLD}} = 4.10 \quad \chi^2_{\text{SLD}} = 1.38
\]
\[
\chi^2_{\text{ALEPH}} = 4.62 \quad \chi^2_{\text{ALEPH}} = 0.34
\]
Apply k-means clustering

Classify clusters by increasing \( \chi^2 \) order in ‘extended’ reduced

\[
\frac{\chi^2}{N_{\text{tot}}} + \sum_{\exp} \frac{\chi^2_{\exp}}{N_{\exp}}
\]

Perform a new sampling with flat priors around the best cluster
PDF results
JAM19 PDFs


\[ Q = 2 \text{ GeV} \]

\[ R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \]

DIS \((p, d)\)

DY \((pp, pd)\)

SIA \((\pi^\pm, K^\pm)\)

SIDIS \((\pi^\pm, K^\pm)\)
JAM19 PDFs


Q = 2 GeV

DIS(p, d)
DY(pp, pd)
SIA(π±, K±)
SIDIS(π±, K±)
JAM19 PDFs


Strong strange suppression

$Q = 2$ GeV

Carlota Andrés
FF results
$Q = m_c$
**JAM19: FF**


\[ Q = m_c \]
JAM19: FF

Q = m_c


Large $\bar{s} \rightarrow K^+$

$zD^{K^+}_{s^+}$
Summary

- **Fitting** several (or all) non-perturbative collinear distributions simultaneously is very challenging

- **MC** statistical methods are important for a robust extraction
  (Crucial for future global TMDs, GPDs analysis)

- **New methodology** needed: **MC multi-steps fit**, k-means clustering, ‘extended’ reduced $\chi^2$

- First simultaneous fit of unpolarized PDFs and FFs: **strange** PDF strongly suppressed
Backup
Impact of SIDIS data
Impact of SIDIS data on PDFs

\[ Q = m_c \]
Impact of SIDIS data on PDFs

\[ Q = m_c \]
Impact of SIDIS data on PDFs

$Q = m_c$

Strong strange suppression
Impact of SIDIS data on FFs

\[ Q = m_c \]
Impact of SIDIS data on FFs

\[ Q = m_c \]
Impact of SIDIS data on FFs

\[ g \rightarrow \pi^+ \]
\[ s^+ \rightarrow K^+ \]

\[ u^+ \rightarrow \pi^+ \]
\[ u^+ \rightarrow K^+ \]

\[ d^+ \rightarrow \pi^+ \]
\[ d^+ \rightarrow K^+ \]

\[ Q = m_c \]

Constraints on
\[ s^+ \rightarrow K^+ \]
Constraints on $R_s$

\[ R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \]

PDFs

$x f(x) + \text{DIS data}$
Constraints on $R_s$

\[ R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \]

$\chi f(x)$

+ DIS data
Constraints on $R_s$

$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}}$$

PDFs

$x f(x)$

$+\text{DIS data}$

$x$  

$R_s$
Constraints on $R_s$

\[ R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \]

PDFs

$xf(x)$

$+$ DIS data

$+$ DY data
Constraints on $R_s$

$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}}$$

PDFs

$x f(x)$

+ DIS data
+ DY data
Constraints on $R_s$

PDFs

\[ R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \]
Constraints on $R_s$

$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}}$

PDFs

$x f(x)$

+ DIS data
+ DY data
+ SIA + SIDIS data
Constraints on $R_s$

PDFs

\[ R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \]

+ DIS data
+ DY data
+ SIA + SIDIS data
Constraints on $R_s$

PDFs

$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}}$$

$xf(x)$

$x$

+ DIS data
+ DY data
+ SIA + SIDIS data
### Chi2

<table>
<thead>
<tr>
<th>Reaction</th>
<th>$N_{dat}$</th>
<th>$\chi^2$</th>
<th>$\chi^2/N_{dat}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIDIS</td>
<td>992</td>
<td>1243.12</td>
<td>1.25</td>
</tr>
<tr>
<td>SIA</td>
<td>444</td>
<td>562.80</td>
<td>1.27</td>
</tr>
<tr>
<td>DIS</td>
<td>2680</td>
<td>3437.96</td>
<td><strong>1.28</strong></td>
</tr>
<tr>
<td>DY</td>
<td>250</td>
<td>416.29</td>
<td>1.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reaction</th>
<th>$N_{dat}$</th>
<th>$\chi^2$</th>
<th>$\chi^2/N_{dat}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIDIS ($\pi^{\pm}$)</td>
<td>498</td>
<td>585.48</td>
<td>1.18</td>
</tr>
<tr>
<td>SIDIS($K^{\pm}$)</td>
<td>494</td>
<td>657.64</td>
<td>1.33</td>
</tr>
<tr>
<td>SIA($\pi^{\pm}$)</td>
<td>231</td>
<td>247.27</td>
<td>1.07</td>
</tr>
<tr>
<td>SIA ($K^{\pm}$)</td>
<td>213</td>
<td>315.53</td>
<td>1.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment</th>
<th>target</th>
<th>hadron</th>
<th>$N_{dat}$</th>
<th>$\chi^2/N_{dat}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPASS</td>
<td>d</td>
<td>$\pi^+$</td>
<td>249</td>
<td>1.26</td>
</tr>
<tr>
<td>COMPASS</td>
<td>d</td>
<td>$\pi^-$</td>
<td>249</td>
<td>1.09</td>
</tr>
<tr>
<td>COMPASS</td>
<td>d</td>
<td>$K^+$</td>
<td>247</td>
<td>1.24</td>
</tr>
<tr>
<td>COMPASS</td>
<td>d</td>
<td>$K^-$</td>
<td>247</td>
<td>1.43</td>
</tr>
</tbody>
</table>