H1 lepton-jet correlations and ML unfolding

Benjamin Nachman
Lawrence Berkeley National Laboratory
on behalf of the H1 Collaboration

30th Lepton Photon Symposium

January 12, 2022
H1 was one of two multipurpose experiments at HERA.

See Henry’s talk earlier in this session
H1 was one of two multipurpose experiments at HERA.

For this talk: 2006-2007 data, 136 pb\(^{-1}\), 320 GeV

I’ll present a measurement of the electron-jet inbalance.
Why electron-jet imbalance?

Born-level configuration, electron and jet are back-to-back

Typically, jets are studied in the Breit frame, where the Born-level configuration is discarded

However, jet production in the lab frame can be useful for probing Transverse Momentum Dependent (TMD) Parton Distribution Functions (PDFs)

See e.g. Lieu et al. PRL (2019) 192003;
Gutierrez et al. PRL (2018) 162001
Jets at H1

Energy flow algorithm (HFS) combines information from tracker and calorimeters

Neural network-based energy regression

1% jet energy scale uncertainty; 0.5-1% lepton energy scale uncertainty
Energy flow algorithm (HFS) combines information from tracker and calorimeters.

Neural network-based energy regression.

1% jet energy scale uncertainty; 0.5-1% lepton energy scale uncertainty.

Challenge: **unfold multidimensional phase space**
Solution: **use deep learning!**

...can do unbinned, high (and variable-)dimensional unfolding

Challenge: **unfold multidimensional phase space**
Unfold by iterating: OmniFold

Detector-level MC

Data

Particle-level MC

Truth

Step 1: Reweight Sim. to Data

Pull Weights

Step 2: Reweight Gen.

Push Weights

Detector-level Data

Particle-level Truth

Nature

Simulation

Detector-level MC

GEANT

Particle-level MC

RAPGAP
DJANGOH
PYTHIA

Push Weights
Unfold by iterating: OmniFold

Measured

Ideal

Data

Truth

Natural

Synthetic

Simulation

Generation
Unfold by iterating: OmniFold

Detector effects are simulated with Geant3 + H1 custom simulation code.
Unfold by iterating: OmniFold

Our default simulations use RAPGAP and DJANGOH

Unfold by iterating: OmniFold

Natural

Synthetic

Measured

Ideal

Data

Step 2:
Reweight Gen.

Unfold by iterating: OmniFold

Measured

Ideal

Natural

Synthetic

Simulation

Generation

Push Weights

Unfold by iterating: OmniFold

Step 1:
Reweight Sim. to Data

Unfold by iterating: OmniFold

Natural

Measured

Data

Ideal

Iteration 2

Synthetic

Simulation

Pull Weights

Generation

Unfold by iterating: OmniFold

Step 1: Reweight Sim. to Data
Step 2: Reweight Gen.

Iteration 2

Unfold by iterating: OmniFold

Iteration 2

Measured

Ideal

Natural

Data

Synthetic

Simulation

Generation

Push Weights

Unfold by iterating: OmniFold

Iterate $n$ times

Step 1: Reweight Sim. to Data

Step 2: Reweight Gen.
Unfold by iterating: OmniFold

Natural

Measured

Data

Ideal

Truth

Synthetic

Simulation

Generation

Unfold by iterating: OmniFold

OmniFold is:
- **Unbinned**
- **Maximum likelihood**
- **Full phase space (compute observables post-facto)**
- **Improves the resolution from auxiliary features**
Unfold by iterating: OmniFold

OmniFold is:
- **Unbinned**
- **Maximum likelihood**
- **Full phase space (compute observables post-facto)**
- **Improves the resolution from auxiliary features**

In this measurement: simultaneously unfold lepton and jet kinematics and report binned spectra for jet $p_T$, $\Delta \phi$, $q_T/Q$, and jet $\eta$.
Neural networks are naturally unbinned and readily process high-dimensional data.
Neural networks are naturally unbinned and readily process high-dimensional data.

We use a trick whereby classifiers can be repurposed as reweighters.
Classification for reweighting

Neural networks are naturally unbinned and readily process high-dimensional data.

We use a trick whereby classifiers can be repurposed as reweighters.

\[
\frac{p_1(x)}{p_0(x)} \approx \frac{\text{NN}(x)}{1 - \text{NN}(x)}
\]

Classifier (NN) trained to distinguish data sampled from \( p_1 \) versus \( p_0 \).
Neural networks are naturally unbinned and readily process high-dimensional data.

We use a trick whereby classifiers can be repurposed as reweighters.

N.B. the distribution is binned for illustration, but the reweighting is unbinned.
Classification for reweighting

All of these distributions are simultaneously reweighted!
Unfold by iterating: OmniFold

Step 1: Reweight Sim. to Data

Detector-level MC

Particle-level MC

Push Weights

Push Weights

Pull Weights

RAPGAP

DJANGOH

PYTHIA

GEANT
OmniFolding $ep$ simulations

We see excellent closure for the full phase space!
Measurement of lepton-jet correlation in deep-inelastic scattering with the H1 detector using machine learning for unfolding

H1 Collaboration

(To be submitted to Physical Review Letters)

(Dated: August 30, 2021)

The first measurement of lepton-jet momentum imbalance and azimuthal correlation in lepton-proton scattering at high momentum transfer is presented. These data, taken with the H1 detector at HERA, are corrected for detector effects using an unbinned machine learning algorithm (OmniFold), which considers eight observables simultaneously in this first application. The unfolded cross sections are compared to calculations performed within the context of collinear or transverse-momentum-dependent (TMD) factorization in Quantum Chromodynamics (QCD) as well as Monte Carlo event generators. The measurement probes a wide range of QCD phenomena, including TMD parton distribution functions and their evolution with energy in so far unexplored kinematic regions.
Results

Table IV. Numerical data on normalized inclusive jet cross sections $1/\text{jet}$ as a function of the lepton-jet azimuthal angular difference. Further details are specified in Table I.

Figure 3. The uncertainty breakdown per observable.
Excellent agreement with fixed order at high $q_T$, excellent agreement with TMD prediction at low $q_T$. 
Parton shower Monte Carlo programs also provide excellent agreement with the data across the spectra.
Simultaneous for free! (binning is for illustration)

Results

**H1**
Stat. Uncertainty

- $Q^2 > 150 \text{ GeV}^2$
- $0.2 < y < 0.7$
- $p_T^{\text{jet}} > 10 \text{ GeV}$
- $k_T, R = 1.0$

<table>
<thead>
<tr>
<th>$q_T/Q$</th>
<th>$\Delta \phi$</th>
<th>$\eta^{\text{jet}}$</th>
<th>$p_T^{\text{jet}}$</th>
</tr>
</thead>
</table>

Correlation

-1.0
0.0
0.5
1.0
Looking forward

Publishing unbinned measurements is tricky - we have started a conversation about this in a recent paper. Feedback is most welcome!

2109.13243
Today, I have presented the first ML-based unfolding with collider data.

This is the start of an exciting program to advance our study of QCD into higher dimensions.

This particular measurement has important constraining power for TMD PDFs and provides important input to planning and design for the future EIC.
Backup
Unfold by iterating: OmniFold

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>Ideal</td>
</tr>
<tr>
<td>Simulation</td>
<td>Measured</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measured</th>
<th>Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>1</td>
<td>100%</td>
</tr>
</tbody>
</table>

Ideal

Natural

Truth

Simulation

Generation
Unfold by iterating: OmniFold

- Ideal
- Measured
- Synthetic
- Natural
- Data
- Truth
- Generation
- Simulation

<table>
<thead>
<tr>
<th>Measured</th>
<th>2</th>
<th>0%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

from 1
from 2

from 1
from 2
Unfold by iterating: OmniFold

- **Generation**
  - Simulation

- **Ideal**
  - Measured

---

<table>
<thead>
<tr>
<th>Measured</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100% 50%</td>
<td>0% 50%</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Natural</th>
<th>Synthetic</th>
<th>Ideal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Simulation</td>
<td>Truth</td>
</tr>
</tbody>
</table>

---

Unfold by iterating: OmniFold

1. **Generation**
   - Synthetic
   - Natural

2. **Measured**
   - 100% 50%

3. **Ideal**
   - 100%

**Axes:**
- x2
- x2/3
- 1
- 2

**Legend:**
- Generation
- Simulation
- Ideal
- Measured

**Notes:**
Unfold by iterating: OmniFold

After iteration 1

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>Simulation</td>
</tr>
<tr>
<td>Ideal</td>
<td>Measured</td>
</tr>
</tbody>
</table>

Measured

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Ideal

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>Data</td>
</tr>
<tr>
<td>Synthetic</td>
<td>Simulation</td>
</tr>
<tr>
<td>Generation</td>
<td>Truth</td>
</tr>
</tbody>
</table>
Unfold by iterating: OmniFold

After iteration 1

- Generation
- Simulation

Measured

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Ideal

Natural

Measured

Truth

Synthetic

Simulation

Generation
Unfold by iterating: OmniFold

After iteration 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measured</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Ideal

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Ideal

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Truth

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Generation

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Unfold by iterating: OmniFold

After iteration 1

<table>
<thead>
<tr>
<th></th>
<th>Generation</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x3/4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>x3/2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ideal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ideal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
</tr>
</tbody>
</table>

Natural

Measured

Synthetic

Ideal

Truth

Data

Simulation

Generation
Unfold by iterating: OmniFold

After iteration 1

Measured
2 0% 50%
1 100% 50%

Ideal

1 2

Natural

Measured

Ideal

Simulation

Truth

Synthetic

Generation
Unfold by iterating: OmniFold

After iteration 2

Measured

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Ideal

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Natural

Data

Truth

Synthetic

Simulation

Generation

Ideal

Measured
Unfold by iterating: OmniFold

After iteration $\infty$

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>Simulation</td>
</tr>
<tr>
<td>Ideal</td>
<td>Measured</td>
</tr>
</tbody>
</table>

N.B. if you just apply $p(\text{ideal} | \text{measured})$, you would have gotten the wrong answer!

<table>
<thead>
<tr>
<th>Measured</th>
<th>2</th>
<th>0%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

| Ideal    | 1 | 2 |
|---|---|
| Synthetic | Simulation |
| Natural   | Truth    