Multidimensional Deconvolution with Profiling

Huanbiao Zhu[1] (huanbiaz@andrew.cmu.edu), Krish Desai[2,3], Mikael Kuusela[1], Vinicius Mikuni[4], Benjamin Nachman[3,5], Larry Wasserman[1]

[1]Department of Statistics and Data Science, Carnegie Mellon University
[2]Department of Physics, University of California, Berkeley
[3]Physics Division, Lawrence Berkeley National Laboratory
[4]National Energy Research Scientific Computing Center, Berkeley Lab
[5]Berkeley Institute for Data Science, University of California, Berkeley

Motivation

- Instrumental effects distort spectra from their true values. Statistically removing these distortions is called **unfolding** in particle physics.
- Modern unfolding methods utilize machine learning to enable unbinned unfolding. **OmniFold[2]** (OF) is among the first classifier-based methods applied to experimental data.
- However, most methods assume the detector response is accurately modeled in simulation, which is only approximately true in practice, with the presence of nuisance parameters.
- This work introduces **Profile OmniFold** (POF), a novel algorithm that extends OmniFold while retaining its key advantages:
  (1) simultaneously profiles nuisance parameters during iteration.
  (2) works with unbinned data.
  (3) utilizes the power of neural network classifiers.
  (4) unfolds multidimensional observables.

Methodology

EM algorithm with the presence of nuisance parameter:

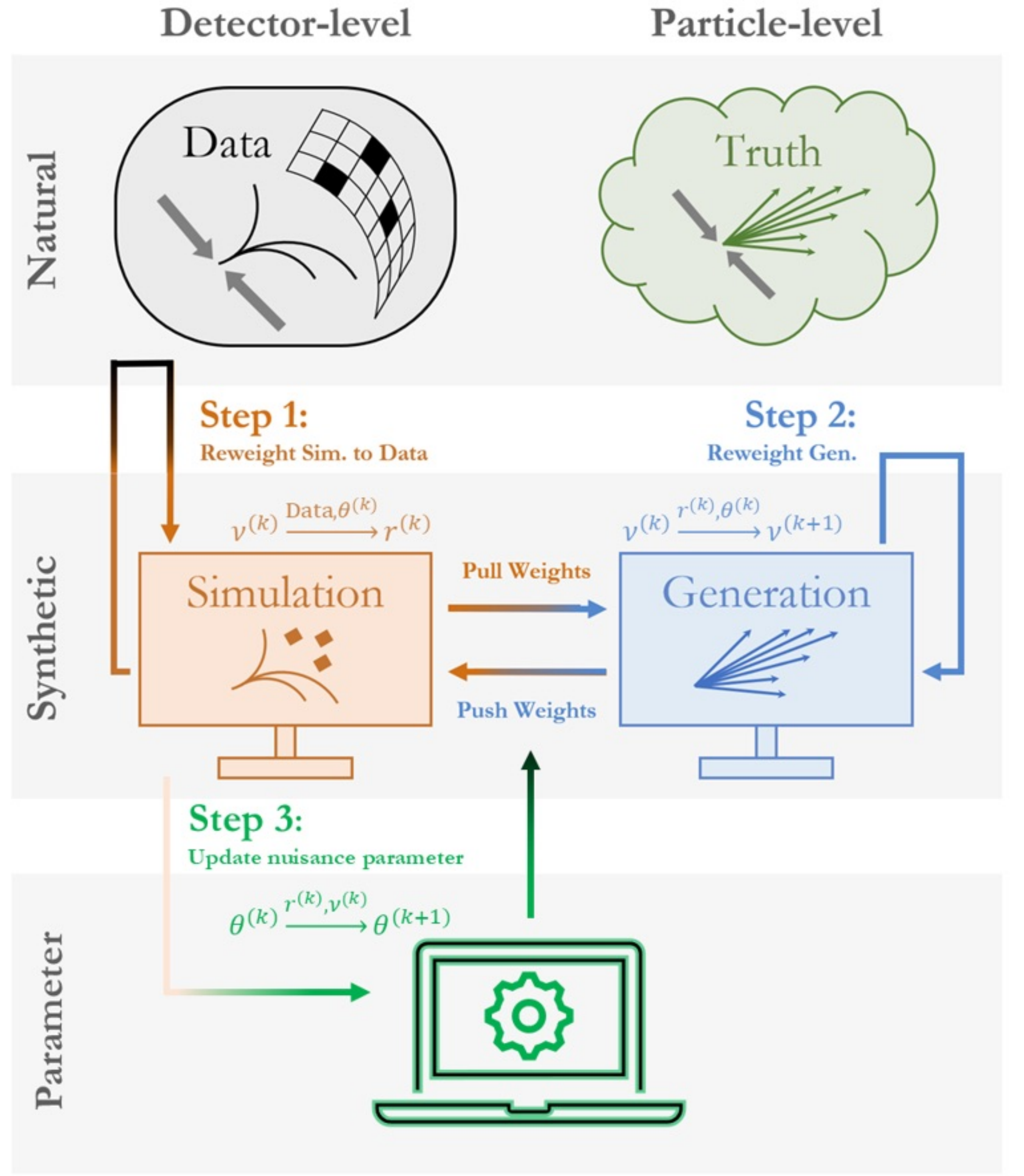
In each iteration, maximize

$$Q(\nu,\theta|\nu^{(k)},\theta^{(k)}) = \int p(y)\int p(x|y,\nu^{(k)},\theta^{(k)})\log p(y,x|\nu,\theta)\,dxdy + \log p_0(\theta)$$
$$\text{subject to } \int \nu(x)q(x)dx = 1,$$

where

ν = reweighting function for the particle-level simulation (theory prediction)
θ = nuisance parameter
x = particle-level quantity
y = detector-level observation
p = experimental distribution
q = Monte Carlo (simulation) distribution
p_0 = prior on θ
w = response kernel reweighting function $\left(w(y,x,\theta) = \frac{p(y|x,\theta)}{q(y|x)}\right)$

Profile OmniFold Algorithm



Profile OmniFold iterates the following steps:

Step 1 reweights the Monte Carlo detector-level simulation to match data.

$$r^{(k)}(y) = \frac{p(y)}{\tilde{q}(y)}, \qquad \tilde{q}(y) = \int w(y,x,\theta^{(k)})\nu^{(k)}(x)q(y,x)dx$$

Step 2 pulls back the detector-level weights to particle level.

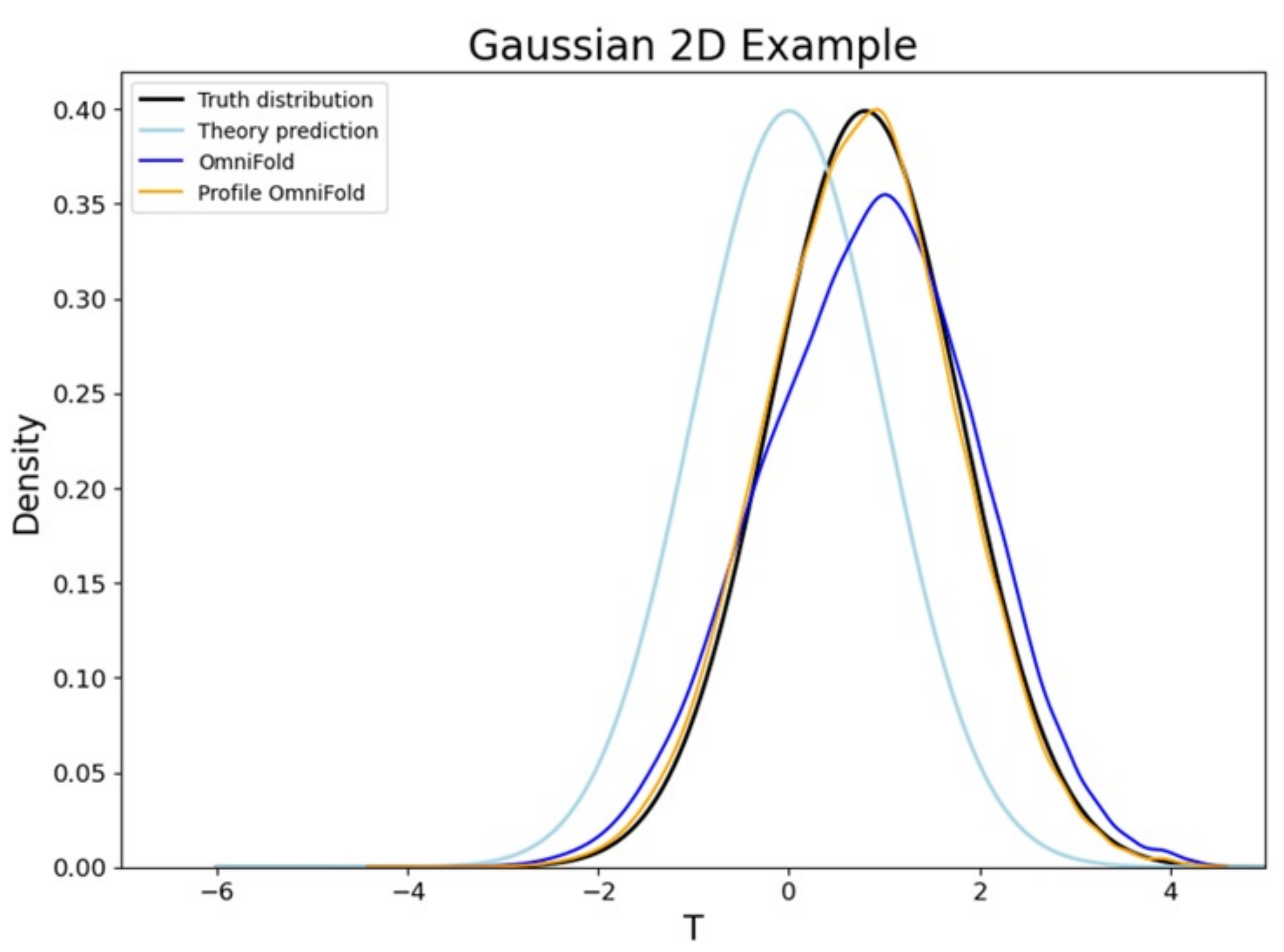
$$\nu^{(k+1)}(x) = \nu^{(k)}(x)\frac{\tilde{q}(x)}{q(x)}, \qquad \tilde{q}(x) = \int w(y,x,\theta^{(k)})r^{(k)}(y)q(y,x)dy$$

Step 3 updates the nuisance parameter based on the detector-level weights and the previous particle-level weights.
Find $\theta^{(k+1)}$ such that

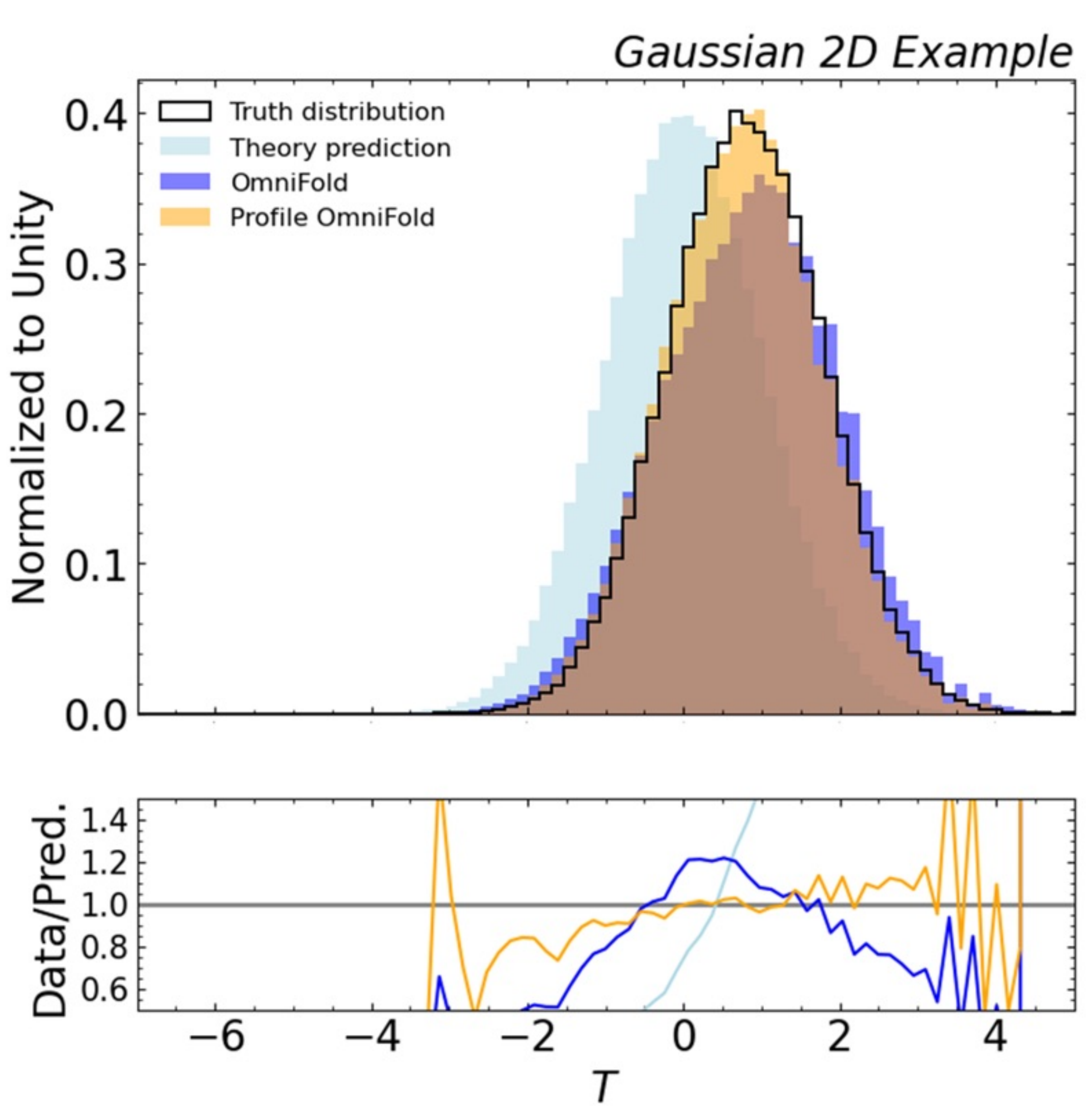
$$\theta^{(k+1)} - \theta_0 = \int\int w(y,x,\theta^{(k)})\nu^{(k)}(x)\frac{\dot{w}(y,x,\theta^{(k+1)})}{w(y,x,\theta^{(k+1)})}r^{(k)}(y)q(y,x)dxdy$$

Experimental Results

$$X \sim N(\mu,\sigma^2),\ Z_1 \sim N(0,1),\ Z_2 \sim N(0,\theta^2)$$
$$Y_1 = X + Z_1, \qquad Y_2 = X + Z_2$$



Results of unfolding 2D Gaussian example. Above: The unfolded particle-level density using POF (orange) and OF (blue) with both algorithms running for 5 iterations. Upper-right: Unfolded spectrum aggregated into 80 bins. Bottom-right: Ratio of the unfolded spectrum to the truth spectrum.



References

[1]Zhu et al. (2024). "Multidimensional Deconvolution with Profiling." In arXiv: https://arxiv.org/abs/2409.10421

[2]Andreassen et al. (2020). "OmniFold: A Method to Simultaneously Unfold All Observables." In Physics Reiview Letter. 124.