

France-Berkeley PHYSTAT Conference on Unfolding

Monday, 10 June 2024 - Thursday, 13 June 2024

Book of Abstracts

Contents

Welcome and Logistics	1
HEP overview (30'+15')	1
Statistics overview (30'+15')	1
ML overview (30'+15')	1
Binned ML methods overview (20'+20')	1
Unbinned Discriminative ML methods overview (20'+20')	1
Unbinned Generative ML methods overview (20'+20')	1
Performance / benchmarking with regularisation choice (20'+20')	2
Unfolding in direct BSM search-sensitive regions of phase space (20'+20')	2
Contributed talks	2
HEP summary	2
A 24-dimensional Cross-section Measurement with ATLAS (15+15)	2
Multidimensional cross-section measurements with H1 (15+15)	2
Closing of the meeting and summary	3
Welcome reception	3
Full Event Particle-Level Unfolding with Variable Length Latent Variational Diffusion (20+20)	3
Response Matrix Estimation in Unfolding Differential Cross Sections (20+20)	3
Dealing with Uncertainties (20+20)	4
QUnfold: Quantum Annealing for Distributions Unfolding in High-Energy Physics (20+20)	4
Simplified Template Cross Sections (STXS) (20+20)	4
Open discussion on challenges and possible solutions in unfolding	5
Unfolding using Denoising Diffusion (20+20)	5

Unfolding in the context of $g-2$ (20+20)	5
Unfolding in the context of a heavy ion analysis (20+20)	5
Unfolding is not unsmearing (20+20)	6
Moment Unfolding (20+20)	6
Profile likelihood unfolding with large number of bins (20+20)	6
Using unfolded data in global QCD analyses (20+20)	7
Likelihood-based unfolding with the CMS Higgs combination tool (20+20)	7

1

Welcome and Logistics

Corresponding Authors: abutter@lpnhe.in2p3.fr, benjamin.philip.nachman@cern.ch, lydia.brenner@cern.ch

2

HEP overview (30'+15')

Corresponding Author: philippe.gras@cern.ch

3

Statistics overview (30'+15')

Corresponding Author: mikael.kuusela@cern.ch

4

ML overview (30'+15')

Corresponding Author: tilman.plehn@cern.ch

5

Binned ML methods overview (20'+20')

Corresponding Author: jingjing.pan@cern.ch

6

Unbinned Discriminative ML methods overview (20'+20')

Corresponding Author: mariel.pettee@gmail.com

7

Unbinned Generative ML methods overview (20'+20')

8

Performance / benchmarking with regularisation choice (20'+20')

Corresponding Author: lydia.brenner@cern.ch

9

Unfolding in direct BSM search-sensitive regions of phase space (20'+20')

Corresponding Author: sarah.louise.williams@cern.ch

10

Contributed talks

11

HEP summary

12

A 24-dimensional Cross-section Measurement with ATLAS (15+15)

Corresponding Author: mariel.pettee@gmail.com

13

Multidimensional cross-section measurements with H1 (15+15)

Corresponding Author: fernando_tta@berkeley.edu

14

Closing of the meeting and summary

Corresponding Author: benjamin.philip.nachman@cern.ch

15

Welcome reception

16

Full Event Particle-Level Unfolding with Variable Length Latent Variational Diffusion (20+20)

Authors: Aishik Ghosh¹; Alexander Shmakov¹; Daniel Whiteson¹; Kevin Thomas Greif¹; Michael James Fenton¹

¹ University of California Irvine (US)

Corresponding Authors: michael.james.fenton@cern.ch, aishik.ghosh@cern.ch, kgreif@uci.edu, alexander.shmakov@cern.ch, daniel.whiteson@cern.ch

Collisions at the Large Hadron Collider (LHC) provide information about the values of parameters in theories of fundamental physics. Extracting measurements of these parameters requires accounting for effects introduced by the particle detector used to observe the collisions. The typical approach is to use a high-fidelity simulation of the detector to generate synthetic datasets that can then be compared directly with experimental data. However, these simulations are often proprietary and computationally expensive. An alternative approach, *unfolding*, statistically adjusts the experimental data for detector effects. Traditional unfolding algorithms require binning data in a small set of pre-selected dimensions. Recent methods using generative machine learning models have shown promise for performing un-binned unfolding in high dimensions, allowing later computation of many observables. However, all current generative approaches are limited to unfolding a fixed set of observables, making them unable to perform *full-event* unfolding in the variable dimensional environment of collider data. A novel modification to the variational latent diffusion model (VLD) approach to generative unfolding is presented, which allows for unfolding of high- and variable-dimensional feature spaces. The performance of this method is evaluated in the context of semi-leptonic $t\bar{t}$ production at the LHC. Additionally, the dependence of the unfolding on the training data prior is assessed by evaluating the model on datasets with alternative priors.

17

Response Matrix Estimation in Unfolding Differential Cross Sections (20+20)

Authors: Larry Wasserman¹; Mikael Kuusela²; Richard Zhu¹

¹ *Carnegie Mellon University*

² *Carnegie Mellon University (US)*

Corresponding Authors: larry@cmu.edu, huanbiaz@andrew.cmu.edu, mikael.kuusela@cern.ch

In unfolding problem, the response matrix is the forward operator which models the detector response. In practice, the response matrix is not known analytically. Instead, it needs to be estimated using Monte Carlo simulation, which introduces statistical uncertainty into the unfolding procedure. This raises the question of how to estimate the response matrix in a sensible way. In most analyses at the LHC, this is done by binning the events and counting the corresponding numbers of events from bins to bins. However, this approach can suffer from undersmoothing, especially with a small sample size. To address this issue, we propose a two-step approach to response matrix estimation. First, we estimate the response kernel on the unbinned space. Second, we propagate the estimated response kernel into an integral equation to obtain an estimate for the response matrix.

18

Dealing with Uncertainties (20+20)

Corresponding Author: kyle.james.read.cormier@cern.ch

19

QUnfold: Quantum Annealing for Distributions Unfolding in High-Energy Physics (20+20)

Authors: Gianluca Bianco¹; Simone Gasperini¹

¹ *Universita e INFN, Bologna (IT)*

Corresponding Authors: simone.gasperini4@unibo.it, gianluca.bianco@cern.ch

In High-Energy Physics (HEP) experiments, each measurement apparatus exhibit a unique signature in terms of detection efficiency, resolution, and geometric acceptance. The overall effect is that the distribution of each observable measured in a given physical process could be smeared and biased. Unfolding is the statistical technique employed to correct for this distortion and restore the original distribution. This process is essential to make effective comparisons between the outcomes obtained from different experiments and the theoretical predictions.

The emerging technology of Quantum Computing represents an enticing opportunity to enhance the unfolding performance and potentially yield more accurate results.

This work introduces QUnfold, a simple Python module designed to address the unfolding challenge by harnessing the capabilities of quantum annealing. In particular, the regularized log-likelihood minimization formulation of the unfolding problem is translated to a Quantum Unconstrained Binary Optimization (QUBO) problem, solvable by using quantum annealing systems. The algorithm is validated on a simulated sample of particles collisions data generated combining the Madgraph Monte Carlo event generator and the Delphes simulation software to model the detector response. A variety of fundamental kinematic distributions are unfolded and the results are compared with conventional unfolding algorithms commonly adopted in precision measurements at the Large Hadron Collider (LHC) at CERN.

The implementation of the quantum unfolding model relies on the D-Wave Ocean software and the algorithm is run by heuristic classical solvers as well as the physical D-Wave Advantage quantum annealer boasting 5000+ qubits.

20

Simplified Template Cross Sections (STXS) (20+20)

Corresponding Author: rahul.balasubramanian@cern.ch

21

Open discussion on challenges and possible solutions in unfolding

Corresponding Author: bogdan.malaescu@cern.ch

22

Unfolding using Denoising Diffusion (20+20)

Author: Camila Pazos¹

Co-authors: Martin Klassen¹; Pierre-Hugues Beauchemin¹; Shuchin Aeron²; Taritree Wongjirad²; Vincent Alexander Croft³

¹ *Tufts University (US)*

² *Tufts University*

³ *Nikhef National institute for subatomic physics (NL)*

Corresponding Authors: vincent.croft@cern.ch, shuchin.aeron@tufts.edu, c.pazos@cern.ch, hugo.beauchemin@tufts.edu, martin.klassen@cern.ch, taritree.wongjirad@tufts.edu

Unfolding detector distortions in experimental data is critical for enabling precision measurements in high-energy physics (HEP). However, traditional unfolding methods face challenges in scalability, flexibility, and dependence on simulations. We introduce a novel unfolding approach using conditional denoising diffusion probabilistic models (cDDPM). By modeling the conditional probability density between detector-level observations and truth-level particle properties from various physics processes, the cDDPM unfolding performance generalizes across varied simulated processes and kinematic distributions without retraining. We demonstrate proof-of-concept on toy models and evaluate on simulated Large Hadron Collider jets across different physics processes.

23

Unfolding in the context of g-2 (20+20)

Corresponding Author: lellouch@cpt.univ-mrs.fr

24

Unfolding in the context of a heavy ion analysis (20+20)

Author: Molly Park¹

¹ *Massachusetts Inst. of Technology (US)*

Corresponding Author: mitay@mit.edu

In this study, the process of unfolding is studied in the context of a heavy ion photon-tagged jet analysis. The SVD and D'Agostini unfolding algorithms are compared, and an application of using the MSE to choose the regularization strength is shown. Additionally, the investigation looks into the bias associated with unfolding in relation to prior choice. The performance is evaluated with different theoretical models and the bottom line test.

25

Unfolding is not unsmearing (20+20)

Author: Michael Schmelling¹

¹ *Max Planck Society (DE)*

Corresponding Author: michael.schmelling@mpi-hd.mpg.de

In particle physics unfolding methods are employed when the basis used to represent an estimate of the truth is not the basis with statistically independent expansion coefficients. For the discrete unfolding problem the latter is given by the eigenvectors of the Fisher information matrix, which measures the amount of information carried by the data about the truth. In typical cases it is ill-conditioned, with the consequence that the measurements constrain only a small number of the expansion coefficients. This allows for highly efficient data reduction, but only for a biased estimate of the truth. Unfolding methods differ in how they bias the result. A way to quantify this is the posterior response matrix.

26

Moment Unfolding (20+20)

Corresponding Author: krish.desai@lbl.gov

27

Profile likelihood unfolding with large number of bins (20+20)

Author: David Walter¹

Co-authors: Josh Bendavid²; Matteo Defranchis¹

¹ *CERN*

² *Massachusetts Inst. of Technology (US)*

Corresponding Authors: david.walter@cern.ch, matteo.defranchis@cern.ch, joshbendavid@gmail.com

In this talk we will discuss previous measurements using binned maximum likelihood unfolding focusing on analyses with large numbers of bins and nuisance parameters. We will outline the technical implementation and highlight the challenges and limitations. Furthermore, we showcase the newly-developed approach of “linearized binned likelihood unfolding”, a modified formalism that has a better scaling and allows to perform unfolding on even larger numbers of bins and nuisance parameters.

28

Using unfolded data in global QCD analyses (20+20)**Corresponding Author:** oleksandr.zenaiev@cern.ch

29

Likelihood-based unfolding with the CMS Higgs combination tool (20+20)**Author:** Alessandro Tarabini¹¹ *ETH Zurich (CH)***Corresponding Author:** alessandro.tarabini@cern.ch

The CMS Higgs combination tool is the software package used for statistical analyses by the CMS Collaboration. The package, originally designed to perform searches for a Higgs boson and the combined analysis of those searches, has evolved to become the statistical analysis tool presently used in the majority of measurements and searches performed by the CMS Collaboration. Since Combine has access to the full likelihood function, it can also be used to perform a likelihood-based unfolding. This approach has become the standard unfolding procedure for many analyses within the collaboration.