The MODE Collaboration Effort

Jan Kieseler

3.3.2022

- selected topics -
Machine-Learning Optimised Design of Experiments

https://mode-collaboration.github.io

At University of Padova Dr. Tommaso Dorigo, Dr. Pablo De Castro Manzano, Dr. Lukas Layer, Dr. Giles Strong, Dr. Mia Tosi, and Dr. Hevjin Yarar
At Université catholique de Louvain Dr. Andrea Giammanco, Prof. Christophe Delaere, Mr. Maxime Lagrange, and Dr. Pietro Vischia
At Université Clermont Auvergne, Prof. Julien Donini, and Mr. Federico Nardi
At the Higher School of Economics of Moscow, Prof. Andrey Ustyuzhanin, Dr. Alexey Boldyrev, Dr. Denis Derkach, and Dr. Fedor Ratnikov
At the Instituto de Física de Cantabria, Dr. Pablo Martínez Ruíz del Árbol
At CERN Dr. Lukas Heinrich and Dr. Jan Kieseler
At University of Oxford Dr. Atilim Gunes Baydin
At New York University Prof. Kyle Cranmer
At Université de Liège Prof. Gilles Louppe
At GSI Dr. Anastasios Belias
At Rutgers University Dr. Claudius Krause
At Uppsala Universitet Prof. Christian Glaser
At TU-München Mr. Max Lamparth
At Durham University Dr. Patrick Stowell
At Lebanese University Prof. Haitham Zaraket
What are we looking for?

- Dark matter
- Stability of the universe
- Grand unifying theories
- Resolution ~ energy
### Underlying Model

**Remarkable theory**
- spanning order of magnitude
- precision predictions

**Stochastic in nature**

**Go deeper: new, more powerful colliders**
Detect Particles and Processes

- We measure decay products of the stochastic process
- Detectors employ parts of or all of the above concepts
Simulation of detector response

- High fidelity simulation of particles interacting with matter
- Carefully validated
- Validity also spans orders of magnitude

http://arxiv.org/abs/2108.02803
• Very high dimensional parameter space
• An expert group at each step
  ‣ Very well-understood steps
  ‣ Use surrogates and short cuts to the final objective (physics result)
  ‣ The chain has **almost no parts** expressed in a differentiable way (or code)
• **Tuning often by hand** and optimisations take a lot of time and person power
• Determining multiple parameters ~ fitting a function

• The optimisation of a detector or a reconstruction chain is conceptually the same thing

To perform this optimisation we need to know $\frac{\partial \epsilon}{\partial \theta}$: how does our photon efficiency change w.r.t. the reconstruction parameters $\theta$?

• Gradients can be calculated
  ‣ Numerically: unfeasible for many parameters
  ‣ Algorithmically: requires unbroken gradients throughout the whole chain: every step needs to be differentiable
    ➤ Differential programming
Interlude: differentiable programming in a

- Differentiable programming is used by, but is independent of machine learning.

- At its core: in any operation, include a way to access its gradient w.r.t. all parameters (if it exists): auto differentiation

- Auto-differentiation is neither pure numeric nor pure symbolic differentiation
  - Numerical differentiation is not feasible for large optimisation problems
  - Fully symbolic differentiation can easily become not feasible from computational point of view

- In most cases, back propagation is used
  \[ \frac{dL}{dA} = \frac{dx}{dA} \frac{dL}{dx} \]
  - A → x → L
  - Calculate the numerical values of \(b = \frac{dL}{dx}\) using the analytic gradient of the operation
  - Calculate the numerical values of \(b \cdot \frac{dx}{dA}\) in the same way

- Each 'atomic' simple operation only needs to be equipped with a simple analytic gradient, then evaluated numerically: best of both worlds.

- This is implemented in one way or another in all modern ML frameworks (TF, torch..)
- Even expressing non-ML algorithms in differentiable frameworks comes with huge advantages w.r.t. the capability of optimising their parameters, and using state-of-the-art libraries to do so

For a nice overview see Atılım Günes Baydin et al (2018), arXiv:1502.05767v4
Use differentiable programming to optimise particle physics detectors given a quantification of the physics target(s) and the detector cost
What do we already have

- Proofs of concept (or more) exist
- Overarching connection is missing
- Next: will be going backwards
A typical high-energy physics analysis

Data
- Includes the full chain up to final state particle reconstruction

Simulation
- Simulate different physics processes
- Simulate detector response and electronics
- Proceed as for data
- Calibrate and correct using control samples in data
- Assign uncertainties on normalisation and shapes

Analysis
- Create histograms of data and simulation
- A contribution of a process or a shape to the data: signal
- Contributions of other processes: background

- Perform profile likelihood fit of parameters of interest and parameters representing variations w.r.t. uncertainties

Uncertainties can be (significantly) reduced by choosing good observables to fit

CMS, arXiv:1603.02303
Automatising analyses on simulation

- Standard analysis concept: perform fit to data based on histograms
  - Best signal fraction (S/S+B)
  - Often use a trained classifier here
  - Lowest uncertainties

- Can take into account uncertainties in to learn not only best S/S+B, but also lowest uncertainty: INFERNO [2]

- Derive “best histogram” and perform a standard profile likelihood fit to extract the physics result

✓ Concepts for generalisable, differentiable analyses workflows exist

Other methods overview in Bremer, Cranmer et al, arXiv:1911.01429
• Determine final state particles and their properties from detector hits

The usual chain
• Local seeding (pattern recognition)
• Local clustering (pattern recognition)
• Software compensation (pattern recognition)
• Identification (pattern recognition)
• Linking of individual detector parts (pattern recognition)

Always the same patterns

Many steps cutting / segmenting information: a priori non differentiable
Use DNNs as generic reconstruction

The usual chain
• Local seeding (pattern recognition)
• Local clustering (pattern recognition)
• Software compensation (pattern recognition)
• Identification (pattern recognition)
• Linking of individual detector parts (pattern recognition)

• Use ML for the task: CNN structure

  ▶ Adapts itself to grid-like granularity
  ▶ Re-optimisation == a few GPU hours vs. months of optimisations by hand

• Compare different segmentations
  ▶ Saturation effects visible

C. Neubüser, JK, P Lujan, arxiv:2101.08150, EPJC
Use known sub detectors in a new way

- At future very high energy colliders
  - Muons will deposit more energy
  - Muons will bent less → tracks provide less information

- The pattern of radiation deposits contains information useful to regress the true muon energy, opens up new possibilities and impacts detector concepts
- Based on CNNs
Irregular Geometries

- Detectors are not grids of sensors
- The reconstruction needs to account for that
- Graph neural networks are a powerful solution
  - No sorting required
  - No grid
  - Sense of connection
  - Basic principle: information exchange through edges (connections)

A typical HEP detector has $O(500k)$ active sensors each event
The network needs to fit into the resources

GravNet

- Developed to overcome resource limitation
- Main ‘trick’: split into low dimensional coordinate and high dimensional feature space
- Tested on a HEP calorimeter reconstruction task
- Up to 2 orders of magnitude improvement w.r.t. resources: ~500k hits can be processed

GravNet in torch_geometric!

Fused kernels
https://github.com/cms-pepr/pytorch_cmspepr

CMS DP-2020/001
Multi-particle reconstruction

The usual chain

- Local seeding (pattern recognition)
- Local clustering (pattern recognition)
- Software compensation (pattern recognition)
- Identification (pattern recognition)
- Linking of individual detector parts (pattern recognition)

Challenges

- A priori unknown number of particles to reconstruct
- Particles are not dense objects with clean centres and boundaries
- The input data is represented by point clouds

N. Wang et al, arXiv:1904.01355
• Maximum number of objects per image/point cloud: number of pixels/vertices

• Learn to move pixels towards the object center

• Map to Gaussian probability

\[
\phi_k(e_i) = \exp\left(-\frac{\|e_i - C_k\|^2}{2\sigma_k^2}\right)
\]

• Assign seed score

\[
\mathcal{L}_{\text{seed}} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{s_i \in S_k\}} \|s_i - \phi_k(e_i)\|^2 + 1_{\{s_i \notin S_k\}} \|s_i - 0\|^2
\]

• Collect (from highest seeds score) around the seeds

• ‘Only’ performs segmentation

• Heavily relies on the center of an object

  › Problematic concept for particles

D. Neven et al, arXiv:1906.11109
B. Zhang, P. Wonka, arXiv:1912.00145
• Merge object property determination and segmentation
  ‣ Created by potentials
• Create a decoupled ‘clustering’ space
• Assign ‘condensation score’ → charge
  ‣ Highest score condensation points carry object properties
• Push non-differentiable ‘clustering’ step towards the very end

**Segmentation**

\[
\hat{V}_k(x) = \|x - x_\alpha\|^2 q_{\alpha k}, \text{ and} \\
\hat{V}_k(x) = \max(0, 1 - \|x - x_\alpha\|) q_{\alpha k}.
\]

**Reconstruction efficiency and noise**

\[
L_\beta = \frac{1}{K} \sum_k (1 - \beta_{\alpha k}) + s_B \frac{1}{N_B} \sum_i n_i \beta_i,
\]

**Object properties**

\[
L_p = \frac{1}{\sum_{i=0}^N (1 - n_i) \arctanh^2 \beta_i} \sum_{i=0}^N L(t_i, p_i) (1 - n_i) \arctanh^2 \beta_i
\]

⇒ Generalises to image data

JK, arxiv:2002.03605, EPJC
High granular calorimeter application

- Differentiable one-shot reconstruction from hits to final state particles
- Can also be applied to information from different sub detectors

The usual chain
- Local seeding (pattern recognition)
- Local clustering (pattern recognition)
- Software compensation (pattern recognition)
- Identification (pattern recognition)
- Linking of individual detector parts (pattern recognition)
What do we already have

Collider / "Space"

Sensors

Geometry

Reconstruction

Analysis

\[ E = [10, 30] \text{ GeV} \]

\[ E = [30, 50] \text{ GeV} \]

\[ E = [50, 70] \text{ GeV} \]

\[ E = [70, 200] \text{ GeV} \]

\[ \sigma_0 \]

\[ \sigma_1 \]

\[ \sigma_2 \]

\[ \sigma_3 \]

\[ \sigma_4 \]

\[ \sigma_5 \]

\[ \sigma_6 \]

\[ \sigma_7 \]

\[ \sigma_8 \]

\[ \sigma_9 \]

\[ \sigma_{10} \]

\[ \sigma_{11} \]

\[ \sigma_{12} \]

\[ \sigma_{13} \]

\[ \sigma_{14} \]

\[ \sigma_{15} \]

\[ \sigma_{16} \]

\[ \sigma_{17} \]

\[ \sigma_{18} \]

\[ \sigma_{19} \]

\[ \sigma_{20} \]

\[ \sigma_{21} \]

\[ \sigma_{22} \]

\[ \sigma_{23} \]

\[ \sigma_{24} \]

\[ \sigma_{25} \]

\[ \sigma_{26} \]

\[ \sigma_{27} \]

\[ \sigma_{28} \]

\[ \sigma_{29} \]

\[ \sigma_{30} \]

\[ \sigma_{31} \]

\[ \sigma_{32} \]

\[ \sigma_{33} \]

\[ \sigma_{34} \]

\[ \sigma_{35} \]

\[ \sigma_{36} \]

\[ \sigma_{37} \]

\[ \sigma_{38} \]

\[ \sigma_{39} \]

\[ \sigma_{40} \]

\[ \sigma_{41} \]

\[ \sigma_{42} \]

\[ \sigma_{43} \]

\[ \sigma_{44} \]

\[ \sigma_{45} \]

\[ \sigma_{46} \]

\[ \sigma_{47} \]

\[ \sigma_{48} \]

\[ \sigma_{49} \]

\[ \sigma_{50} \]

\[ \sigma_{51} \]

\[ \sigma_{52} \]

\[ \sigma_{53} \]

\[ \sigma_{54} \]

\[ \sigma_{55} \]

\[ \sigma_{56} \]

\[ \sigma_{57} \]

\[ \sigma_{58} \]

\[ \sigma_{59} \]

\[ \sigma_{60} \]

\[ \sigma_{61} \]

\[ \sigma_{62} \]

\[ \sigma_{63} \]

\[ \sigma_{64} \]

\[ \sigma_{65} \]

\[ \sigma_{66} \]

\[ \sigma_{67} \]

\[ \sigma_{68} \]

\[ \sigma_{69} \]

\[ \sigma_{70} \]

\[ \sigma_{71} \]

\[ \sigma_{72} \]

\[ \sigma_{73} \]

\[ \sigma_{74} \]

\[ \sigma_{75} \]

\[ \sigma_{76} \]

\[ \sigma_{77} \]

\[ \sigma_{78} \]

\[ \sigma_{79} \]

\[ \sigma_{80} \]

\[ \sigma_{81} \]

\[ \sigma_{82} \]

\[ \sigma_{83} \]

\[ \sigma_{84} \]

\[ \sigma_{85} \]

\[ \sigma_{86} \]

\[ \sigma_{87} \]

\[ \sigma_{88} \]

\[ \sigma_{89} \]

\[ \sigma_{90} \]

\[ \sigma_{91} \]

\[ \sigma_{92} \]

\[ \sigma_{93} \]

\[ \sigma_{94} \]

\[ \sigma_{95} \]

\[ \sigma_{96} \]

\[ \sigma_{97} \]

\[ \sigma_{98} \]

\[ \sigma_{99} \]

\[ \sigma_{100} \]

\[ \sigma_{101} \]

\[ \sigma_{102} \]

\[ \sigma_{103} \]

\[ \sigma_{104} \]

\[ \sigma_{105} \]

\[ \sigma_{106} \]

\[ \sigma_{107} \]

\[ \sigma_{108} \]

\[ \sigma_{109} \]
• Our simulation is highly complex, stochastic, and not differentiable

• Some parameters inherently have no gradient: e.g. adding/removing a sensor layer, switching positions, …
  ▸ There are ideas, and some developments in the direction of solving this
  ▸ This is a very interesting conceptual challenge to contribute to!

• Example: use local surrogates of the gradient (GAN) for optimising the SHIP muon shielding

Sweeping magnet optimisation

- Local differentiable surrogates can help solve the problem of non-differentiable simulation

Figure 7. Muon hits distribution in the detection apparatus (depicted as red contour) obtained by Bayesian optimization (Left) and by L-GSO (Right), showing better distribution. Color represents number of the hits in a bin.

• Atmospheric muons: 1/s/hand area
• Interact only sparsely with material
• Are scattered enough to be used for imaging applications

• Detectors: usually panels, with spatial resolution and detection efficiency
• Optimal starting point for a differentiable simulation

• TOMOPT: package to optimise muon tomography detectors (work in progress)

L. Bonechi, R. D'Alessandro, A. Giammanco
arXiv:1906.03934
The target of MODE is to design and offer to the community a scalable, versatile architecture that can provide end-to-end optimisation of particle detectors, proving it on a number of different applications across different domains.

Study cases:
- **Use known detector concepts in a new way:** demonstration of muon energy measurement in calorimeter
- **Optimise starting from ‘simple’ applications**
  Muon tomography detector optimisation in progress
- **Rethink decades old paradigms:**
  Hybrid calorimeter design integrating tracking layers activity starting

Other use cases being considered / about to start include:
- Hadron therapy
- Muon collider detector shielding

The developed architectures for optimisation are modular
- recycle part of the work for one application when moving to the next one
- Very happy about any suggestions / contributions

---

Dear Colleagues,

Initiated by the European Committees for Astroparticle (APPEC), Particle (ECFA) and Nuclear Physics (NuPECC), and following a first joint seminar held in Oraev in 2019, Expressions of Interest for common activities have meanwhile been endorsed in the following areas:
- Dark Matter (iDMEu)
- Machine-learning Optimized Design of Experiments (MODE)

Jan Kieseler