



Autodiff Status & Plans

Lino Gerlach¹

¹Princeton University

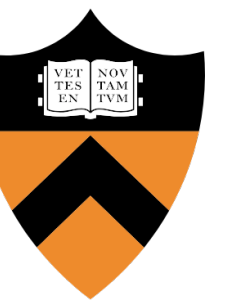
IRIS-HEP Demo Day

1 November 2024



Introduction

- About me: started postdoc @ Princeton this summer
 - 75% on CMS trigger / 25% on IRIS-HEP (more like 90-10 so far)
- Identified autodiff as potential field to contribute @ phydep.dev
- Potentially high impact topic, but little activity recently
- This talk: make people aware of what I plan to do, discuss general direction. Report (very small) progress. Just getting started, so I am open to tips & suggestions!



Autodiff in HEP

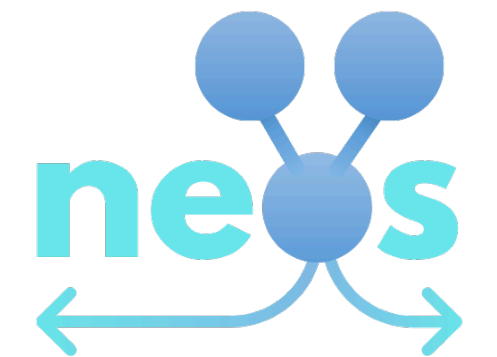
- Autodiff activities in IRIS-HEP mainly under [gradhep GitHub organization](#)

- No activity for over a year (on main branches)



- HSF had a 'Differentiable Computing' activity

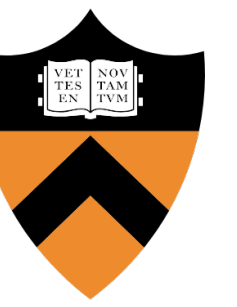
relaxed 😴



- Not active anymore, moved to archive for reference



- Why? Lack of person power? Road blocks / show stoppers?



Recent work on AD in IRIS-HEP

- Most recent project (afaik) by Saransh Chopra (we met at phyhep.dev)
- Enable autodiff w/ jax for awkward arrays ([slides](#))
- Final result: awkward, coffea, Vector are now differentiable
- What's missing: Realistic use case to check for limitations & bugs
- Ideal test case: Fully differentiable version of AGC (my plan)

Awkward
Array





Towards AD for the AGC

- Github repo towards fully differentiable AGC: [agc-autodiff](#)
- Plan for my next steps:
 - Work towards differentiable AGC
 - Report any problems / showstoppers encountered on the way

Prepared small live demo on this

alexander-held commented on Jul 31, 2023 • edited

Following the [jet calibration example](#) discussed at [pyhep.dev](#), there are a few different ways forward in the direction of extending functionality. Some of these can be pursued in parallel.

The `pipeline` function contains them all:

```
def pipeline(jets, a):  
    """analysis pipeline: calculate mean of dijet masses"""  
    return np.mean(get_mass(correct_jets(jets, a)))
```

- replace `correct_jets` by more complex calibration operations, e.g. with a differentiable `correctionlib`
- replace `get_mass` by a more complicated physics analysis, e.g. involving combinations of objects and sorting (the AGC top mass reconstruction is a good candidate)
- replace `np.mean` by statistical model construction + inference

In addition to this:

- use `nanoevents` -> [feat: example with nanoevents and jax #2](#)



New Standalone AD Library?

- Jax used for most differentiable HEP approaches. Some issues:
 - Quite large, cumbersome integration w/ awkward, incompatible w/ awkward's JIT
- Jim suggested to develop a pure auto-diff library (e.g. w/ pybind11 & numba)
 - Full control over package, compatibility w/ awkward, small & easy-to-install
 - Would other packages (e.g. neos & relaxed) be interested, too? (Or just for awkward?)
- Potential project for a summer fellow next year?



Live Demo: Differentiable Trijet Top Mass Reconstruction