Extending Automatic Differentiation Support in RooFit

...

Abhigyan Acherjee

About Me

BS+MS, Computer Science, University of Cincinnati

Previous Experience:

- Worked on data simulation and comparisons for LHCb.
- Worked on improving an optimization algorithm (Minuit) in maximum likelihood fits
- Achieved a 38% reduction in runtime for certain use cases.

Overview of Project

Extend Automatic Differentiation Support in RooFit:

Automatic differentiation gives an edge over numerical differentiation and more accuracy and scalability for problems with large number of parameters, as opposed to using a likelihood function written by hand in C++.

There have been various ongoing efforts into modernizing RooFit. Further proof of work is needed on production grade workflows.

Current Goals:

- Extend the AD support in RooFit by transpiling more RooFit primitives.
- Profile and extend if necessary the underlying minimization infrastructure in the minuit package.

Very Broad Overview of Minuit Investigation

Item 1	Item 2	Item 3
1.Delineated a test suite that encapsulated behaviors often observed in Minuit.	2. As the complexity of the fit (error) function increased, the accuracy of each parameter fitted decreased.	3. Using parameter fit values from a subset(epoch) of the data as the starting values for fitting the entire data on >=8-parameter functions, we were able to reduce the runtime of the fit by 38%.

Specific Goals

- Develop an understanding of the current codebase and experiment offering speedups through Automatic Differentiation.
- Investigate the speedups offered through Minuit and Minuit optimisation strategies.
- Develop functionalities and classes to extend support to HistFactory and other parts of RooFit
- Profile the benchmarks and understand performance bottlenecks of the current AD-aware tutorials.
- Explore opportunities for parallelism using OpenMP or CUDA.

Timeline Part 1:

Week 1 (14th-20th Aug) :Develop an understanding of the current codebase and experiment offering speedups through Automatic Differentiation.

Week 2 (21st-28th Aug) :Examine the AD-aware tutorials of RooFit. Enumerate a list of tutorials which need to support AD.

Week 3 (28th Aug-3rd Sept) : Investigate the status of AD support in the HistFactory benchmarking models.

Week 4(4th Sept -10th Sept) : Improve the benchmark infrastructure with focus on the AD use case by adding more tests.

Week 6,7 () (18th Sept - 1st Oct) :Profile the benchmarks and understand performance bottlenecks of the current AD-aware tutorials.

Week 8(2nd Oct - 8th Oct) : Develop functionalities and classes to extend support to HistFactory and other parts of RooFit. For example, support Higgs to gamma-gamma analysis (via the opendata website).

Week 9 (9th Oct-15th Oct) :Extend the Higgs to a more complex example for the Higgs to gamma-gamma analysis in rootbench.

Timeline Part 2:

Week 11,12(23rd Oct-5th Nov) Implement the fit from the IRIS HEP analysis grand challenge (AGC) in RooFit with the AD backend, implement missing support the needed RooFit primitives for that usecase. Benchmark and compare with other AGC implementations.

Week 13,14 (6th Nov-19th Nov) Incorporate the AGC benchmarks in the public benchmark infrastructure.

Week 15,16 (20th Nov-3rd Dec): Support a tutorial that contains numeric integrals by using Clad's fallback mode to numerically differentiate only such integrals and use AD for the rest of the model.

Week 17,18 (8th Jan-21st Jan) Investigate improving the Minuit seeding procedure. Deliverable: initialize the minimizer without having to numerically calculate the second derivatives for all parameters"

Week 20 (29th Jan-5th Feb) :Prepare documentation and tutorials to provide support for the classes in the code squashing for AD.

Week 21, 22 (5th Feb-18th Feb) :Explore opportunities for parallelism using Open MP/CUDA.

Week 23 (19th Feb-25th Feb) Develop final documentation and tutorials. Deliverable: a document on how to write new RooFit primitive support; Technical document on the usage of AD in RooFit at scale.

Week 24 (26th Feb-4th March) : Prepare a final presentation.

Questions?