



Contribution ID: 119

Type: Oral

Automatic differentiation of binned likelihoods with RooFit and Clad

Wednesday 26 October 2022 12:30 (20 minutes)

RooFit is a toolkit for statistical modeling and fitting used by most experiments in particle physics. Just as data sets from next-generation experiments grow, processing requirements for physics analysis become more computationally demanding, necessitating performance optimizations for RooFit. One possibility to speed-up minimization and add stability is the use of automatic differentiation (AD). Unlike for numerical differentiation, the computation cost scales linearly with the number of parameters, making AD particularly appealing for statistical models with many parameters. In this talk, we report on one possible way to implement AD in RooFit. Our approach is to add a facility to generate C++ code for a full RooFit model automatically. Unlike the original RooFit model, this generated code is free of virtual function calls and other RooFit-specific overhead. In particular, this code is then used to produce the gradient automatically with Clad. Clad is a source transformation AD tool implemented as a plugin to the clang compiler, which automatically generates the derivative code for input C++ functions. We show results demonstrating the improvements observed when applying this code generation strategy to HistFactory and other commonly used RooFit models. HistFactory is the subcomponent of RooFit that implements binned likelihood models with probability densities based on histogram templates. These models frequently have a very large number of free parameters, and are thus an interesting first target for AD support in RooFit.

References

Experiment context, if any

Significance

This contribution will demonstrate significant advancements in state-of-the-art capabilities for both High-Energy Physics (HEP) and computer science domains. The widespread use of automatic differentiation of statistical models in HEP will significantly improve the performance and numeric stability of statistical analysis. On the computer science side, this work demonstrates that source-transformation based automatic differentiation can be added to complex libraries like RooFit. It also showcases an application where different AD strategies can be compared, as other research groups are experimenting with other AD implementations for differentiable likelihoods (usually the ones available in the Python ecosystem, such as TensorFlow).

Primary authors: SINGH, Garima (Princeton University (US)); REMBSER, Jonas (CERN); MONETA, Lorenzo (CERN); VASILEV, Vassil (Princeton University (US))

Presenters: SINGH, Garima (Princeton University (US)); VASILEV, Vassil (Princeton University (US))

Session Classification: Track 2: Data Analysis - Algorithms and Tools

Track Classification: Track 2: Data Analysis - Algorithms and Tools