

Fifth MODE Workshop on Differentiable Programming for Experiment Design



Contribution ID: 122

Type: Talk

Bringing Automatic Differentiation to CUDA with Compiler-Based Source Transformations

Tuesday 10 June 2025 12:30 (25 minutes)

GPUs have become increasingly popular for their ability to perform parallel operations efficiently, driving interest in General-Purpose GPU Programming. Scientific computing, in particular, stands to benefit greatly from these capabilities. However, parallel programming systems such as CUDA introduce challenges for code transformation tools due to their reliance on low-level hardware management primitives. These challenges make implementing automatic differentiation (AD) for parallel systems particularly complex.

CUDA is being widely adopted as an accelerator technology in many scientific algorithms from machine learning to physics simulations. Enabling AD for such codes builds a new valuable capability necessary for advancing scientific computing.

Clad is an LLVM/Clang plugin for automatic differentiation that performs source-to-source transformation by traversing the compiler's internal high-level data structures, and generates a function capable of computing derivatives of a given function at compile time. In this talk, we explore how we recently extended Clad to support GPU kernels and functions, as well as kernel launches and CUDA host functions. We will discuss the underlying techniques and real-world applications in scientific computing. Finally, we will examine current limitations and potential future directions for GPU-accelerated differentiation.

Authors: KOUTSOU, Christina (Princeton University (US)); LANGE, David (Princeton University (US)); Mr VASSILEV, Vassil (Princeton University (US))

Presenter: KOUTSOU, Christina (Princeton University (US))

Session Classification: Methods and tools

Track Classification: Methods and tools