Exploring Autodiff and Finite Differences Using the Rosenbrock Function

Nicholas Rubinó, Adrian Celaya, David T. Fuentes Ph.D
Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX

Background

- Computer models are growing increasingly complex as more accurate models are created. Optimizing parameters in these models relies on efficiently taking derivatives.
- Using existing machine learning frameworks, we can compute derivatives of physics-based functions quickly.
- Our model is used to find the optimal flip angles for hyperpolarized magnetic resonance imaging (HPMRI).

Using the Rosenbrock function as an example to see the benefit of this approach as a surrogate for a full physics-based model in HPMRI.

Hypothesis

Using automatic differentiation to recover the parameters of the extended Rosenbrock function will be faster than numerical differentiation.

Methods

- The Rosenbrock function is used as a standard reference frequently used to evaluate the efficiency and accuracy of automatic differentiation frameworks.
- The function consists of two variables and is of the form:
  \[ f(x, y) = (1-x)^2 + (x - y^2)^2. \]

Numerical Differentiation (Sum of Finite Differences)

- Uses the chain rule to break down functions into elementary operations using dummy variables to store how each piece interacts with the others.
- Has problems with truncation error as well as rounding error due to the imprecise nature of finite point arithmetic with numbers very close to zero.
- Cost of computation as well as rounding errors exacerbated when computing a gradient.
- Requires \(O(n)\) evaluations for \(n\) dimensional gradient.

What is Automatic Differentiation (Autodiff)?

- A technique used to find the gradient of a function.
- Uses the chain rule and evaluates the gradient to each basic operation in the forward pass
- The final gradient is then computed by multiplying the nodes together from the derivative trace.
- For 2, 4, and 10 parameters, finite differences was faster than autodiff. However, for all parameters, auto reverse completed the objective function optimization with fewer function evaluations. Thus auto reverse is the best candidate for the HPMRI optimization function shown in Figure 5.

Reverse Mode

- Starts with a forward pass where a trace of the primal nodes of the function is computed and each node is augmented with adjoint nodes, which store all intermediate variables and their connections in memory.
- This is called a Wengert list.
- Adjoints are denoted by a variable with a bar on top of it.
- The derivatives of each adjoint are not computed or stored during the forward pass.
- Then there is a reverse pass where the Wengert list is followed and the derivatives of each adjoint node are computed.
- The derivatives of each adjoint are chosen with respect to a given dependent variable.

Discussion/Conclusion

Auto reverse was the fastest of the three methods. Auto forward and reverse took the same number of function evaluations but auto reverse was much slower overall. We have not been able to get any of these solvers to work for the HPML function yet due to the high complexity of the function. We expect auto reverse to be the best option.

References

- Cohen, William. Automatic Reverse-Mode Differentiation. 1-3 Bayesian Signal model for HPMRI