

High Order Reverse Mode in Automatic Differentiation

Mu Wang

Advisor: Alex Pothen

April 17, 2017

PhD Thesis Abstract:

We study the high order reverse mode of Automatic Differentiation (AD) in this dissertation. Automatic Differentiation (AD) is a technique to augment a computer program for computing a function so that the augmented program computes the derivatives as well as the function values. AD is employed to solve optimization problems and differential equations from many application domains, and has been included among the top twenty algorithms in scientific computing. The reverse mode of AD propagates the derivatives in the reverse order of the evaluation of the objective function. It has optimal time complexity for computing first-order derivatives since it implements the Baur-Strassen theorem, which states that the complexity of evaluating all (first order) partial derivatives of a scalar objective function is only a constant factor greater than the complexity of computing the objective function itself.

We propose the generalized high order reverse mode of which the first order reverse mode can be considered as a special case. The key concept in the high order reverse mode is live variables. The set of live variables at a step is defined as currently active variables whose values will be used in future steps of the computation. Then the invariant of reverse mode AD can be stated as follows: The intermediate results at a step k in the algorithm are the derivatives of a suitably defined equivalent function $f_k(S_k)$, where S_k constitutes the current set of live variables, and f_k is obtained by composing the elemental functions by which the function is computed. A general expression for the high order chain rule for evaluating derivatives of $f_k(S_k)$ from values at the previous step, $f_{k+1}(S_{k+1})$, yields the high order reverse mode. We have provided a thorough complexity analysis of these algorithms.

The algorithms are implemented to exploit both sparsity and symmetry. Sparsity is exploited by performing updates with only the nonzero values in the algorithm. Symmetry is exploited by keeping only the unique elements

in the high order derivative tensor due to the inherently high degree of symmetry. For second and third order derivatives, we show that enabling pre-accumulation can further reduce the time complexity.

The combinatorial properties of the high order reverse mode are also discussed in this dissertation. We prove that the second order reverse mode is equivalent to a combinatorial model that performs vertex elimination on the computational graph of the gradient. More generally, performing vertex elimination in a specified order on the computational graph corresponding to the procedure for evaluating derivatives up to the d -th order yields the $(d+1)$ -th order reverse mode.