How automatic differentiation improves numerical weather forecasting and the potential for gradient based machine learning in future improvements

Daniel Holdaway[*]
Global Modeling and Assimilation Office
NASA Goddard Space Flight Center `dan.holdaway@nasa.gov`

Abstract

When it comes to forecasting of the latest hurricane or severe weather event there is often spirited public discussion about who produced the best forecast: the European model or the American model. The European model often wins out, forecasting the eventual track of a particular storm a few days before. There are of course many differences between these models but one of the key differences lies in the use of the forecast model adjoint in the European model.

Numerical weather forecasting consists of a complex physical model and a data assimilation system. The physical model, though highly skillful, drifts from reality due to inaccuracies in the model physics as well as the initial conditions. Periodically the model must be constrained to observations using the data assimilation system. Typical data assimilation systems are based on Bayes theorem and provide an algorithm for finding the most probable state given the most recent forecast and set of observations. Typically state of the art models consist of around one billion degrees of freedom and a new initial condition must be generated within around twenty minutes to maximize the observation gathering window. Usually around 50 million per day are assimilated. Given the complexity of the problem certain assumptions are made by the assimilation system so that a timely solution can be sought. These include that the probability distributions are Gaussian and that the model behaves linearly over short time ranges. Once this is taken into account a cost function is formulated and a gradient descent is used to find the optimal solution. Because of the time varying nature of the observations, the need for the gradient of the cost function and the assumption of linearity the adjoint of the model is a term in the algorithm.

To be fair to the American weather forecast modeling community it should be noted that the European model really has one job: provide a global medium time range weather forecast. In the US we have many different models providing global weather forecasts, regional forecasts e.g. for Tornado Alley, forecasts for forest fires, high resolution simulations of hurricanes, seasonal forecasts, climate forecasts, etc so resources are more spread. A benefit the Europeans enjoy is that they are able to manually generate and maintain the adjoint of their model. Given the range of applications a hand written adjoint system would not be well suited to the American community, instead the model gradient is represented using an ensemble of model states. Recently it is becoming clear that this approach is sub-optimal. An alternative, that can also work in the context of a diverse range of modeling applications, is to use automatic differentiation. In the past this has not been deemed to be a sensible approach since the resulting code is not

efficient enough and because automatic differentiation tools are not designed to take advantage of the iterative nature of data assimilation, resulting in superfluous forward sweeps of the code.

In this paper we show that these issues can be overcome. We have generated the adjoint the finite volume cubed sphere dynamical core used in the NASA Goddard Earth Observing System using the Tapenade automatic differentiation tool. The code can be generated from scratch within around a week of development, meaning it can work in the context of a highly collaborative modeling environment. Further we have developed a completely custom interface for the checkpointing that takes advantage of the iterative nature of data assimilation. With this interface the checkpoints are held memory and testing can be done to ensure checkpoints are not superfluous or completely zero, reducing the memory footprint. The resulting adjoint model is shown to be efficient enough for data assimilation applications and scientifically superior to using the ensemble.

Gradient-based machine learning is beginning to gain traction as a useful tool for numerical weather prediction and data assimilation. We will outline some of these applications and discuss the progress that has been made so far. We also speculate on a number of future potentials of the method.