Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-140-RC1, 2019

Geoscientific Model Development Discussions

Open Access

Interactive comment on "Algorithmic Differentiation for Cloud Schemes" *by* Manuel Baumgartner et al.

Anonymous Referee #1

Received and published: 12 July 2019

This paper discusses the application of Algorithmic Differentiation (AD) to the simulation of cloud schemes. The results are interesting and very well discussed and interpreted. Therefore, I think that the paper represents a valuable contribution for the targeted journal. However, the presentation could be improved at several places as detailed below.

p. 1, l. 5: "to spot parameters" sounds a bit too colloquial to me, I would prefer "to identify parameters"

p. 2, l. 12-13: "it is largely unused in meteorological contexts" I really see the paper as an important contribution. However, there are numerous papers discussing AD in the context of meteorology, already the platform www.autodiff.org lists over 50 contributions. The development of the AD tool TAF was motivated by climate simulations.

Therefore, the above statement is a bit misleading. I would say that the potential of AD is not fully exploited or something a long this line.

p. 3, l. 8: "this representation" The reference of "this" is not clear, since the previous sentence already describes the application of the chain rule. In my opinion, "this representation" should refer to the original code interpreted statement by statement. In the current formulation this is not the case.

p. 3, l. 24-25: Strictly speaking your function has the 4 inputs a,b,c,d Therefore, you also have to provide 4 dot values (\dot{a}, \dot{b}, \dot{c},\dot{d}) and not only (\dot{a}, \dot{b}).

p. 3, l. 27: "all common elemental operations ... are indeed differentiable" Not quite right, this would exclude sqrt, abs, max, \, ... AD theory assumes that these functions are evaluated at a point such that the elemental operations are differentiable in a neighbourhood of the argument. This allows the elemental operations mentioned before but excludes the evaluation at critical points.

p. 4, l. 17-18: "In AD Lingo" sounds too colloquial to me, reformulate

p. 4, l. 19-20: "v = t_1*t_2" I would use here "w = t_1*t_2" to have the same notation as on page 3 for this example. This also effects Eq. (6)

p. 6, l. 16: Here one could cite the corresponding paper

Andreas Griewank, Kshitij Kulshreshtha, Andrea Walther On the numerical stability of algorithmic differentiation Article in Computing, Springer Vienna, 2012

showing the stated accuracy

p. 14, l. 11-25: The discussion about the dependence on the step size is the only critical issue for me

I think this is completely misleading. If the time step is larger, the simulation is done for a larger time interval, therefore it is somehow reasonable that the sensitivity increases.

Looking at one time step you do not compute the sensitivity at a specific point in time but of function that is defined over a whole time interval. Therefore the scaled sensitivities are completely fine and reasonable. Therefore, the discussion is misleading and should be corrected. Furthermore, there are several papers dealing the the time step issue and the accuracy of the derivatives provided by AD. To name a few:

Peter Eberhard, Christian H. Bischof Automatic Differentiation of Numerical Integration Algorithms Article in Mathematics of Computation, 1999

Adrian Sandu On the Properties of Runge-Kutta Discrete Adjoints Conference proceeding, Computational Science – ICCS 2006, Springer, 2006

Andrea Walther Automatic differentiation of explicit Runge-Kutta methods for optimal control Article in Journal of Computational Optimization and Applications, 2007

Mihai Alexe, Adrian Sandu Forward and adjoint sensitivity analysis with continuous explicit Runge-Kutta schemes Article in Applied Mathematics and Computation, 2009

Some of them could be mentioned

p. 19, l. 11-15: discussion on the computational effort for the forward mode I think that it does not become clear that one computes ONE directional derivative for twice the operations to evaluate the function. Furthermore, if one wants to compute a whole gradient using the forward mode, one should use the vector forward mode which leads to a significant reduction in complexity. This should be mentioned here.

p. 19, l. 19-28: As already described above I do not agree with these statements. Some of the papers mention above analyse also the convergence behaviour for the integration methods obtained by AD for the sensitivity equation and the adjoint equation. Since you are using the standard 4th order Runge Kutta method, the derivatives computed with AD approximate the analytical solution of the sensitivity equation and the adjoint equation, respectively, with the same order as the standard 4th order Runge Kutta method for the state equation.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-140, 2019.