

Interactive comment on “An optimized treatment for algorithmic differentiation of an important glaciological fixed-point problem” by Daniel Goldberg et al.

B. Christianson (Referee)

B.Christianson@herts.ac.uk

Received and published: 11 March 2016

General comment

I really like this paper.

It's a nice application of an established fixed-point iteration method to a new area, explained in a way that should facilitate use of the approach elsewhere. The iterator used is interesting, because it is sufficiently contractive to converge in a relatively small number of iterations (unlike those used in CFD, for example) but is not super-linear (unlike, say, Newton constructors), and so adjoint iteration is required and the choice of adjoint start-point is potentially significant.

C1

There are two audiences for this paper: geoscientists looking to apply AD efficiently to their specific problems; and those already familiar with AD wanting to apply this particular approach in another application domain. It may be worthwhile to insert additional references to the standard AD literature in order to help the second group get the most out of this paper; and to help the first group learn more about AD, which in turn will give them access to techniques which originated in other application domains.

I make some more specific comments, and suggestions for changes, below: I stress that these really are just suggestions.

Specific comments

line 87: the mechanical adjoint was originally proposed by J.C. Gilbert, "Automatic Differentiation and Iterative Processes", *Optimization Methods and Software* 1(1) (1992), 13-22, and it might be useful to cite the discussion in Gilbert's paper, as well as that in Christianson 1994. That the mechanical adjoint doesn't always actually solve the adjoint fixed-point problem accurately - or at all - was pointed out by Gilbert: the quick test whether it did is to check if the adjoints corresponding to u_0 are close to zero, where u_0 is the starting value for the forward iterations.

line 155: This would be a good point to insert some references to standard AD literature: as well as the excellent book by Griewank and Walther already in the Reference list, there is a brisk introductory survey paper (available open-access) by Bartholomew-Biggs et al, "Automatic Differentiation of Algorithms", *JCAM* 124(1-2) (2000), 171-190.

line 164: this observation remains true even when the matrix is not self-adjoint.

line 193: it doesn't have to be the Euclidean norm, contraction with respect to any operator norm will do!

line 202: "uses a fixed point loop to calculate (7)" - not quite. The fixed point loop in Christianson (1994) deliberately calculates (10) rather than (7). This is for two reasons: to avoid repeatedly adding numerically (very) small terms to big ones; and in order to

C2

allow a “warm start” by using an “arbitrary” initial value of $\delta^* w$ that is close to the fixed point. It may be worth moving equation (10) earlier in the paper and pointing out explicitly that : (a) equation (7) converges with n to the value of $\delta^* a$ -hat that corresponds to the fixed point of equation (10) ; (b) equation (10) converges to the correct fixed point regardless of what starting point for $\delta^* w$ is actually used ; and (a) equation (7) corresponds to the result of calculating $\delta^* a$ -hat after iterating equation (10) precisely n times starting from $\delta^* w = \delta^* u$; Table 4 seems to assume starting at $\delta^* w = \delta^* u$, but there is no need for this restriction.

line 245: see above discussion on line 202.

line 250: it would be nice to know what norm is being used for the forward convergence: logically the adjoint norm should be used for the reverse convergence. (For example, the sup norm should be used in reverse if the 1-norm is used forwards; the euclidean norm is self-adjoint, etc.) To first order, the error in the calculated value of the cost function J is the inner product of the error in u (from the forward pass) with the converged value of $\delta^* w$. This inner product is bounded by any vector norm of the error in u multiplied by the corresponding adjoint vector norm of $\delta^* w$: Christianson, “Reverse accumulation and implicit functions”, Optimization Methods and Software, 9 (4) (1998), 307-322.

line 393: the point about indirect solvers being more efficient in large dimensions is a good one, but (as well as having the best derivative values) the final forward iteration also generally has the best pre-conditioner.

Technical comments

line 54: applying the chain rule to the numerical values line 76: correspond to a discretization of the correct line 123: of a nonlinear operator F to obtain u : line 165: - this analytic approach allows invocation line 198: required to ensure convergence of Φ to a fixed point line 229: undone at the end of each iteration. Once convergence is reached, storing takes place as normal in the POSTLOOP phase. line 232: simplest

C3

to replace certain specific sections of OpenAD-transformed code line 253: would not require changes to this subroutine [obviously it will affect what the subroutine does!] line 285: i’m really not clear why these are uniformly set to zero line 300: presumably $m_{\{i,j\}}^{\{fp\}}$ is the value obtained using BC94? line 330: state the range from Figure 2 explicitly here. It would also be useful to have iteration counts for forward and reverse convergence (rather than having to deduce them from Table 5.) line 340: in reverse order relative to forward computation. line 354: recover variable values from the forward computation, so that they can be used in the adjoint computation. line 359: only one level of checkpoints is required. line 442: closer to the forward sweep Fig 1(d) caption: useful to know how the 2nd order differencing was done. Fig 2: seems to have an outlier at 10^{-4} . Any idea why? Table 5: what is the significance of the red and blue entries?

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-11, 2016.

C4