Overview:
For over 2 decades a popular control method has been used in glaciology in order to infer properties of ice sheet bed lubrication and stiffness, and has made use of the adjoint to the stress balance operator in order to find the gradients required, and there is a fairly common approximation to the adjoint calculation. In this work the authors have developed an ice flow model based on a robust finite element code and used automated differentiation to do away with the approximation. They show in the frequency domain the level of retrieval of parameters over a range of settings, and compare the approximate and exact adjoint. Further, they formally investigate how to efficiently calculate the exact adjoint, taking into account such things as the a priori accuracy of the forward model and the regularization of the inverse model.

I think this paper makes a worthy contribution and i would recommend it for publication, moreover there were a lot of things about it that I liked, for instance the discussion of how numerical tolerance and regularization plays into exact adjoint considerations, examination of errors in frequency space, and consideration of sparse data. However, there are issues and errors which need to be fixed before publication.

Broad comments on methodology and content

Writing exact parameters with very localized spectral peaks is nice for the demonstration given. But what happens when more broad spectrums (e.g. Gaussian functions) are considered?

I appreciate what the authors are trying to do with real ice sheet geometry, but the problem is still manufactured since beta is specified again through trig functions, and so this is still synthetic data.

Moreover, to use realistic (measured) velocities may lead to very different results, although the accuracy of inversions could no longer be assessed. But in studies like Morlighem et al 2013 (JGR-earth surface, which, by the way, should be mentioned for comparison), and the last experiment in Goldberg and Sergienko 2011, using satellite-inferred data implies that the difference between exact and approximate adjoints is lessened in terms of lowering j. This does not compare accuracy; **possibly** the full adjoint is picking up short wavelengths more accurately, but is there any way of knowing??? I feel that the tests in this study, with their very simple spectrums of true beta, might not fully address this.

I am not familiar with DassFlow, and I think this is true of most intended readers. Therefore there is relevant information that may be published elsewhere, but is nonetheless not mentioned in this manuscript and should be, since automated differentiation is being discussed. What is the nature of the linear solver, i.e. Is it direct, CG, AMG, or something else? Is is parallel? And – here is why i ask – does Tapenade then parse and differentiate the linear solver subroutines, or are they “bypassed” as in MITgcm (see Heimbach, Hill and Giering 2005 [Fut. Gen. Comp. Sys.], or Goldberg and Heimbach 2013 [TCD, or perhaps now TC])? I think this is fairly important if one is interested in the generation of adjoints, as certain linear solvers, if differentiated by an AD tool, might cause longer adjoint computation times than others. And the authors should also give estimates for the relative computation time of the adjoint and the forward solve, e.g. for accumulation of all forward steps, and for the accumulation from the point where residual = 0.1, the strategy suggested by the authors.

Also, I realize English is not the first language of the authors, but I can tell it is written

“with a French accent” and the instances are too numerous to list. I will try to point out places where the manuscript is particularly difficult to follow.
I am also curious, though this is to the authors to decide – they show that the accuracy of the adjoint can be retained by truncating the accumulation, and they give this truncation in terms of the forward solve. Might the required truncation depend on noise, slip ratio, etc, as does the accuracy of the adjoint? This would be a far more useful result.

Specific Comments

eqs (4-5): point out this is only appropriate for x-z! In (x,y,z), ***t*** is a plane, not a direction, and you would need to project the velocity onto this plane and then apply the sliding law..

3858 l9: not sure what is meant by sequel – companion paper?

3859 l2: interesting, i did not know this – a reference would be nice. But in general i like the discussion in this paragraph.

3862 l1: not sure “control” is the right word

3862 l5: phrasing: test aims to control the partial...

3863 l13: states
l26: on the velocity

3864 l7: comes to retain? Do not understand

3864 say what u_obs is and where it comes from? Becomes clear later but not here (and you never actually plot it anyway)

3865 l26 the forefront?

3867 l3: a few?
L6: i think “a priori” is used incorrectly here.

Section 3 and figs 2-3:
I am very confused as to what delta k is! (as presented in eq 17). i have looked through and cannot find any indication. Is it a single element of the vector? A fourier mode? But I think the result at the end of sec 3 is useful

3868 2nd paragraph: I think i see your point but needs clarification.

3868 l14: sloppy??? also why are you calling this a channel? It is a 2d (x-z) simulation.

Eq (19): say n only goes from 1-3; also use a different symbol as i got confused momentarily with glen's law exponent

3869: say clearly what thickness is (and dx if not already) or clearly direct the reader to the appropriate table

3873 l4: profile **that is** rather convex?

I like the discussion at the end of 4.1 a lot. However the English needs just a little work, i found a few mistakes... here and for the rest of the manuscript nouns are written as singular that should be plural (e.g. Layer**s**)

3875 paragraph at line 21: I don't get what you are trying to say in this paragraph. Discrepancy = misfit?

3876: eq (26) not (19).

3876 paragraph at line 20: is it not 5h and 2h?

3877 l14: did you maybe reverse 6 and 13?

3877 l16: the topography (and ice thickness) are nowhere shown!!!

conclusion: caution, not *pre*caution

3881, l20: huh? Retention of **only** the last 2 states?? this contradicts your finding at the end of section 3!