Comment on gmd-2021-90

Anonymous Referee #2

Referee comment on "fenics_ice 1.0: a framework for quantifying initialization uncertainty for time-dependent ice sheet models" by Conrad P. Koziol et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-90-RC2, 2021

This paper presents a mathematical method and a software tool for quantifying uncertainty in glacier flow model projections. The methods consists of (1) using automatic differentiation to compute the action of the full Hessian of the negative log posterior, (2) computing the smallest few eigenvalue / eigenvector pairs of the Hessian in order to find out which directions in parameter space are most important to sample, and (3) computing the uncertainty in a given output quantity of interest by linearizing the model physics around the most probable state. The authors then test this method on a synthetic test problem, the ISMIP-HOM test case. The authors also address two key shortcomings of how data assimilation in glaciology is usually practiced: accounting for discrete (as opposed to dense) spatial observations and correlation between measurements. Overall, the method is described fairly well albeit with a few points that could use clarifying. I have a few concerns about how generalizable the method is, but in any case the paper and the software are a valuable contribution that I recommend for publication with minor revisions.

The authors make excellent use of low-rank approximations to the Hessian. This trick has appeared in the literature before (although it's not as widely used as it should be). How does their approach compare to that of, say, Petra et al. 2014 or the hIPPYlib code?

I have two real concerns with the approach, although these don't change my overall opinion of this very good paper. First, the method relies on the assumption of quasi-linearity in several places. While the authors check that this assumption wasn't violated for their particular test case, it's difficult to assess whether this would generalize to other problems. By contrast, the stochastic Newton approach in Petra et al. 2014 uses an assumption of quasi-linearity only locally, to bootstrap a more sophisticated Monte Carlo sampling algorithm. Second, the authors make a big deal about using the full Hessian instead of the Gauss-Newton approximation while at the same time relying on quasi-linearity, and yet the Gauss-Newton approximation is works best when the dynamics are almost linear. At the end of the paper, the authors state that they did not use a time-dependent control method because it's computationally expensive. Would a time-

dependent method have been feasible if the authors had instead used the Gauss-Newton approximation? Establishing whether the full Hessian is really necessary is a very important point. Other researchers might want to emulate the techniques described in this paper and yet they might be using other modeling frameworks for which the Gauss-Newton approximation is feasible to implement while the full Hessian is not.

General comments

45, "gradient-based optimisation": The point of this sentence is to state that computing the MAP estimate doesn't give any information about parametric uncertainty. The fact that you used a gradient-based algorithm concerns more the "how" than the "what"; it isn't really important and you could just cut this phrase entirely. You could have used a derivative-free optimization algorithm -- it's a dreadfully awful idea, but you could do it!

57-60: How much better is using the full Hessian than using the Gauss-Newton approximation? This isn't immediately apparent from the text or from the sources you cite here.

97: By taking the parameter-to-observation map f to be a map from R^n to R^m, you're assuming a "discretize, then optimize" mindset. It might clarify the points you make about mesh dependence later on if you instead define it as a map from some function space Q to R^m -- the "optimize, then discretize" mindset. See Gunzburger 2002.

The parameter-to-observation map that you've written down encapsulates both the physics and the measurement process. This is just a suggestion, but it might help the exposition to instead define f as the composition of two maps. First, there's a function g that takes the parameters to the observable fields, like the ice velocity. This function g is basically just "solve the shallow shelf equations". Next, there's a function h that takes the observable fields to the actual observations. When doing the "wrong thing" that you point out later, h is the identity map and the norm is an L^2 norm. When doing the correct that that you have actually implemented, h evaluates the observable fields at a bunch of discrete points and packs these observations into a vector, and the model-data misfit is a discrete sum of squared errors.

135-139: Using a control method does not necessarily imply that you're writing the model-data misfit as a squared L^2 norm, it's just a sinful thing that many glaciologists (including me) have done because it's easier.

It's also worth noting that when you say "mesh dependence", certain readers are immediately going to think of something other than what you describe here. In the PDE-constrained optimization literature, mesh dependence refers to what happens when you use a bad optimization algorithm based on using the vector of coefficients of the derivative

obtained from the adjoint method as a descent direction. This can give really obvious mesh imprinting artifacts in the results, especially with higher-order finite element bases. (The minimal right thing to do is to multiply by the inverse of the mass matrix. You mention using the BFGS method later, and taking the H_0 matrix to be the inverse mass matrix works there as well.) This problem is more a question of how you solve a particular optimization problem. The mesh dependence that you're talking about is much more serious -- by neglecting the discreteness of the observations, going to a finer mesh implicitly assumes that you magically have more measurements than you did before. At a higher level, what you've done is tackle the fact that everyone has been solving the wrong problem, irrespective of how they were solving it. Since some readers will immediately associate with the first case, it might be good to either (1) clarify the distinction with a reference to, say, Schwedes et al. 2017, or, (2) if you don't feel like talking about that, use a different phrase besides "mesh dependence".

140-146: Including the delta term is essentially adding the prior information that you think the mean of the parameters is zero. I don't think this is a good prior in all cases. Are there other ways to get a prior with bounded covariance that don't make this assumption? For example, you could use the Moreau-Yosida regularization of bounds constraints, which instead assumes that the parameters don't wander too far outside a preset interval but which provides no constraints within that interval.

260-264: It's easy to get the impression from this paragraph that you're doing time-dependent inversions, which is only dispelled in the conclusion on line 552. You should probably state this earlier in the text.

270-273: How do you know how many Picard iterations is adequate? Why not use another globalization strategy, like damping / line search or trust regions?

274: Why did you use BFGS when you can calculate a Hessian-vector product? Why not Newton-Krylov?

341: The L-curve is fine, but you might want to mention the discrepancy principle or other more statistically-motivated approaches. See Habermann et al. 2013.

484-490: This is really great, I haven't seen anyone address this issue before.

537-538: I don't think the technical hurdles are minor at all because of exactly what you say in the next sentence.
I would remove this statement from the text.

550-551: It might be worth citing some of the work that Karen Willcox and her group have done on multi-fidelity modeling and UQ.

Technical corrections

Several of the authors' "note to self" comments remain in the manuscript.

484: "isotroptically" -> "isotropically"

References

https://doi.org/10.1137/130934805

https://hippylib.github.io/

https://doi.org/10.1137/1.9780898718720

https://doi.org/10.1007/978-3-319-59483-5

https://doi.org/10.5194/tc-7-1679-2013