Review of « Automatic adjoint-based inversion schemes for geodynamics: Reconstructing the evolution of Earth's mantle in space and time » by Ghelichkhan et al.

The manuscript presents a workflow to realize a geodynamic grail: full inversion of mantle flow together with tectonics to generate digital twins of the Earth. The paradigm in the past to was to push plates at the surface of rheologically simple models and either push them as long as possible so that the last tens of Myrs could become adequate or employ and adjoint for an inversion of the final state given by a thermo-chemical translation of seismic tomography images. Inverting surface kinematics was proposed recently by Li et al. (2007) and Bocher et al. (2016, 2018) on similar types of models (2D). What does this manuscript bring compared to the former studies? It proposes an adaptive framework thanks to automatic differentiation and open software machinery. One could argue for the adjoint method keep physical consistency over the reconstructed duration. The generalization of the approach to other geodynamic problems makes it also very interesting for the community. I find the paper is well presented, provides most of what is needed and giving the most of the needs to evaluate if one wants to dive into the inverse geodynamic adventure.

I have general recommendations for improvements.
 - The most important one is how the Taylor test (or gradient test) is performed. Indeed the residual has to be O(h^2) WHEN h tends to 0. So one can find that it is O(h^2) for some values of h, but it is the most important that it is O(h^2) at the vicinity of machine epsilon. Olivier Talagrand who pioneered the adjoint methods for data assimilation was hard on me on this issue and sent me a variety of gradient test that are successful and failing. Some of them show a residual O(h^2) for large h and become O(h) close to h. So the authors should diminish delta m by orders of magnitude until they reach machine epsilon and the residual becomes then stable. If the residual is O(h^2) close to when it is close to machine epsilon, everything is good. Many papers show this kind of plot. For geodynamics, you can check my preprint on automatic differentiation of StagYY (https://eartharxiv.org/repository/view/6398/ - it can be also interesting to include it in your discussion since we both took parallel paths). By doing so, Figures 4 and 9 will leave no doubt on the efficiency of your method. I you want me to send you the document of Olivier, let me know.
 - The automatic differentiation section (Dolfin). This is very important to be more explicit since it is a new methodology that you take to geodynamics, and this is precisely this methodology that empower your workflow. It is too abstract now. I would prefer you explain more precisely how it works and how it respects the idea that the generated adjoint is the exact adjoint of the forward code. I understand it is hard, since it is why I have such hard time publishing my preprint. But a little more would be nice. Maybe I did not have enough time, but it seems to me that some of the important preparation work to make for automatic differentiation is not explicit. Do you have directives in the forward code? Where do you tape, or checkpoint? What is your organization? It is nice to give a glimpse to the future users of what to be careful about if they want to modify what you have done? Off course, if it's too technical in the paper, that may not work.
 - I would put the derivation of the adjoint equation in an Appendix, you could refer to when you talk about the linear properties of the adjoint with pseudo-plasticity.
 - In my opinion, the discussion needs reworking. For now, most of it reads like a conclusion. The results should be compared to former studies. Bocher et al. (2016, 2018) are ideal because the models have similarities in geometry and objectives. You can discuss then, how the inversion performs. You have all the results for this in your figures. You have to discuss also computing time. This is very important also because you state you want your method to tackle grand challenges of 3D spherical reconstruction with self-generating plates. For now your forward model is 2D, with only 3 orders of magnitude variations in viscosity, which is not enough yet to have plate-like behavior. What is the computing power needed for inversion of this 2D case? To move on, how do you envision going to 6 or more orders of magnitude and 3D? How many million years will you be able to invert for given your computing time? This is a necessary discussion here, since you bring the ambition of the method from the start.

Recommendations along the text:
- Davies and Richards, 1992 in the intro: you should propose a more up to date review here, since so many discoveries have been made since.

- l.36: « state is unknown »: not only. Also because the physics we use is incomplete, and the physical and chemical properties of the Earth remain uncertain.
- l.39: Rolf and Tackley citation here does not feel appropriate. If you want to add a citation with pseudo-plasticity, you souhld choose one focusing on the equations.
- l.43: all the citations here are from the same group, except Dannberg, although other groups have used state-of-the-art 3D convection calculations at high resolution with different codes. Actually, you cite some of them in the next paragraph so you could cite them here too.
- l.44: I would remove the discussion on Monte-Carlo technique. It's not necessary here. It feels out of the scope.
- l.50: Bocher et al. (2016) is a Kalman filter. So it should be cited later. Bocher et al. (2018) is ensemble Kalman filter. Both these methods are doing a similar work as you propose later so they are useful for you to compare. The major difference is that we did not use exactly the same « observations » for the inversion and did apply a Kalman smoother to correct past states, which is something you are interested in.
- l.53. « assimilated » is not a correct word here. I know it has been misused for years. But now you are considering REAL assimilation of surface kinematics, you see the problem. What you call assimilation here is forcing the surface of a model, so it is a forward operation. While assimilation in data assimilation is an inverse operation, specifically the one you are designing here. So please, change it to « nudged » or some more appropriate term.
- Figure 1. In line with the « assimilation » word. Case (a) can be a « nudged » model. It is a way to differentiate nudging (a), KF data assimilation (b) and 4Dvar data assimilation (c).
- L.62: Kalman smoother is a Kalman filter method to do retro-propagate the information in past states. So it is possible to use Kalman filter and improve the knowledge on initial conditions. The major difference with the adjoint is that it is a statistical treatment whereas the adjoint keeps the physical consistency (it has its advantages and disadvantages, considering that the physical model is not perfect).
- l.65: I would remove that idea of an emerging field. We could say that the inverse models of the geoid back in 1990s were inverse geodynamics. Or the pionnering works of Peter and Alik are 20 years old already and we cannot say the technique has penetrated the geodynamics community. Actually, what you propose could help to do so.
- l.74. About Li et al. (2017). It can be important to be nuanced saying that such recovery is limited by the ill-posedness, which is their conclusion.
- Intro: I would say in the intro that solving adjoint equations does not necessarily generate the exact adjoint of the forward code, while AD does. Hence, your approach enforces that the adjoint code is the exact adjoint of the forward code, which is necessary for optimal inversion.
- Table 1: Delta T = Temperature difference between top and bottom ; d= thickness of the system. It is not just characteristic something.
- l.219: it would be nice to cite Talagrand, 1997 or some of his pioneering work here.
- l.225-230: Okay for the advantages. But which drawbacks? --> not the exact adjoint of the forward code which may not be good for inversions. Write a new code for each new problem (set of equations).
- l.231: I think here you could start with AD right away because with AD you find a solution to both the limitations of the continuous adjoint: exact adjoint and versatily of problems (not rewriting the code if you change the problem/update the code). I would reduce the length and you go straight to the point.
- l.233: you can cite Giering and Kaminski, 1998 for AD pioneering work.
- l.240: too bad you forget to cite some existing AD tools used for science application purposes: Tapenade, TAF, or Enzyme.
- l. 240 and equivalence of the methods: I don't agree: the solvers, the methods involve a suite of operations that are discrete anyways and can involve some differences that can possibly prevent the continuous adjoint to be the exact adjoint. It's not only a question of resolution but the approximations you make also, and the tools you use to solve the problem.
- 2.2.2 -> More would be needed here. It's a major point of your work.
- l.560: I find your forward model to be a little oversold here. With 3 orders of magnitude viscosity change, you don't make plates in convection models, and it's even more unlikely to have Earth-like subduction zones (one-sided etc…). Using this model is perfect for the demo here, but I am sure you want to have a more plate-like behavior in the mode in the future, which means necessarily increasing this contrast.
- For the duration of your model, it is worth giving a glimpse of what Earth time it corresponds to, so that we can evaluate how far back in time we can go given the uncertainty on the past state

we look for. Simply putting dimensions here it would be close to 200Myrs. But probably the convection velocities are lower when I look at your figure and it would probably translate to half of this maybe? That would be useful.

- l.612: Visco-plasticity means surface decoupling and more non-linearities. It matters and you could show it by showing the adjoint: the sensitivity of your misfit to surface velocities is probably lower than the isoviscous case. But again, 3 orders of magnitude is not really plate-like so pushing the interpretation is not necessarily good here.
- l.563: surface velocities, except today, are not observed. They come from kinematic models.
- 708: Seton et al., 2023 provides an overview of plate reconstructions that should be discussed here.
- 711: It's not an inversion, but the work of Bello et al. (2015) assesses in some ways the errors produced on the forward modelling.
- 715: The issue with smooth solution is that geodynamic features are sharp and their shape is fundamental for the self-consistency of the flow, especially with variable viscosity. This is very different from seismic or gravity inversions in which smoothing does not hinders the fundamentals of the physics at play.
- l.723: not « assimilation » what you do now with surface velocities is real assimilation :)
- Conclusion can be reduced to the essentials.