Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-363-RC1, 2022

Comment on gmd-2021-363

Anonymous Referee #1

Referee comment on "spyro: a Firedrake-based wave propagation and full waveform inversion finite element solver" by Keith J. Roberts et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-363-RC1, 2022

spyro: a Firedrake-based wave propagation and full waveform inversion finite element solver

Thank you for submitting your manuscript. The paper reads nicely and is well written. The objectives and proposed results are interesting and are a good fit for this journal.

Overall, I would recommend minor revision as some details need to be improved and the presented results are not fully convincing. I think this software could be of benefit to the geophysical community and that an improved manuscript would greatly improve its outreach.

Mainly my concerns are:

- The computational aspect, emphasized in the introduction, needs more clarity and some type of baseline

- The author does not discuss meshing when a "good smooth model" is not available. What would be the strategy if the mesh needs to be modified at each (couple) iteration.

- The 3D results are not very accurate. While I understand these take a lot of resources and can not necessarily be re-run, justification on the accuracy is needed to justify the use of that software.

The author needs to justify better why the software they propose improves the current software landscape for FWI.

Please see below for more details about the manuscript.

Detailed comments

Line 18:

This is a bit terse, other methods exist (NMO, WEMVA, kirchoff, Depth/time extrapolation) and are widely used.

Line 20:

"However, the FWI problem is challenging to apply in practice since there exists a non-unique 20 configuration of data that can best explain the observations. "

This is commonly described as cycle skipping. Please add references to the literature.

Line 47:

One other problematic issue (the main one I would say) that isn't mentioned is the model dependence of the mesh. If the meshing becomes velocity/Unkown dependent then in theory, it should be added to the mathematical definition of the problem and the gradient of the mesh with respect to the velocity must be taken. This problem leads in some cases into elements wrapping onto themselves.

This should be discussed by the author, and considering the mesh independent of the inverse problem needs to be justified.

Figure 1: Did you make that figure? This looks fairly familiar so please add a reference in case this is from somewhere else.

Equation 1: This is not the conventional way to define the wave equation in an acoustic medium. Ignoring the PML for simplicity, the acoustic wave equation (d'Alember operator) is defined as

`diff(u, (t, 2)) - c^2 laplacian(u)`. including the spatially varying wave speed as `div(c ^ 2 grad(u))` is quite unconventional, what is the justification behind this formulation?

Equation 9: Following the point above, the expression for the sensitivity is quite unconventional as well. (usually is either `-2/c^3` or `1` depending on whether the wavespeed or squared slowness `1/c^2` is considered.

Gradient subsampling: This is known to lead to artifacts and inexact gradient. What exact subsampling ratio is used and how much information is lost with that subsampling (maybe compare gradients with and without subsampling).

Line 379: " were not ready to be used in our FWI code." So did you implement the adjoint by hand?

Line 380: Rapid Optimization Library. Any reason you did not choose a standard python library easier to use (i.e scipy) since, as you mentioned in the introduction, the bulk of the computing is the wave equation solves. Therefore "fast" optimization implementation doesn't really impact the overall performance.

Line 386: "we exclusively rely on the second-order optimization method L-BFGS (". L-BFGS rely heavily on exact gradient information. Subsampling the gradient leads to an inexact gradient that doesn't satisfy l-bfgs requirement to guarantee better convergence.

Line 451: "parallelism is trivial and handled by splitting the MPI communicator into groups of processes at initialization" Using MPI for separable task parallelism is known to be overkill and prone to issues. Any reason more adapted task parallelism (Dask, spark, ...) was not used?

The computational runtime, while informative, does not say much about the performance of the proposed method without any reference. How long would it take to run FWI with FD with the same amount of resources? And how much memory?

2D Marmousi

The peak RAM usage seems a bit high. AS a comparison, as standard FD on this model (17km x 3 km discretized on a 20m grid) requires `151*851*1251*4/(1024^3) = .6Gb` of memory for a wavefield sampled at 4ms (`r=4` for your case) which s about 5x-10X lower than what your method seems to require. Can you elaborate on the memory needs of your method?

The runtimes need a reference as by themselves they do not provide any information.

3D overthrust

The 3D overthrust results arent' very convincing as they show a lot of artifacts and noise. I would consider trying to improve those. The author states "which is likely a result of poor source illumination beyond several kilometers of depth.". The illumination is usually corrected for with the hessian (or some diagonal approximation such as `1/norm(u_^2)`. Because you use l-BFGS you should have an approximation of the hessian that is quite accurate after that many iterations. Therefore this interpretation seems incorrect. You can see in related work (such as Witte et. al you refer to) that the deep part of the model can be recovered with a god initial model such as yours.

A 5Hz peak wavelet is unrealistic as it leads to <1Hz data not being feasible in real life. Low-frequency FWI is usually performed for frequencies >3Hz. This could be achieved with a higher frequency wavelet (ie 15 Hz) than band/lowpass your data to 3-8Hz for inversion.