We thank for taking the time to review the manuscript as the topical editor in great detail. We think that incorporating the critic greatly improves the quality of the manuscript and makes it easier to understand.

All upcoming line and page references refer to the new manuscript.

Dear authors,

Thank you for submitting your manuscript to GMD. Given the challenges we had to find suitable reviewers for your work, I decided to provide a more substantial Topical Editor comment to replace the missing second review.

The paper investigates solutions to the Stokes equations with applications to ice flow modelling. The Stokes equations are discretised using the finite-element method using FENICS. Two solvers implement the Picard and the Newton method, respectively. The authors apply their numerical solver to community benchmarks and discuss convergence rates as function of different step size selection strategies.

Although being an interesting and highly relevant topic, the presented work is insufficient in several aspects, mostly in terms of relevance and quality standards within GMD. In summary, the work seems to be a stripped down version of an already published preprint, where formal mathematical proofs were removed (https://arxiv.org/abs/2307.02930). Also, the work lacks in providing significant geoscientific context with respect to the applications. From the succinct text, it is difficult to grasp the novelty of the proposed work, given that the fact that Newton is faster than Picard is exactly what one would expect for the p-Stokes problem. Although it is a nonlinear problem, the p-Stokes equations are the optimality conditions of a strictly convex functional. As a result, one would expect Newton's method to work very well.

Due to the critic, we added experiments and changed the title a bit. We think that the novelty is the exact step size, which improves the Picard iteration and Newton's method. Especially, the Picard iteration with exact step sizes seems to be a good choice compared to Newton's method as it reliably needs nearly the same number of iterations for every step (Fig. 20, Fig. 22, Table 2 and Table 4).

The presented results are also somewhat strange. The poor convergence of Newton solvers in most of the cases are suspicious and could reflect an implementation issue in the algorithm. It would also be interesting to see on more iteration at which "relative difference" (or error) the Picard solver stalls. Such stalls can in most cases originate from either poor scaling of the error, or an issue with the code. In general, the convergence rate of Newton's method should outperform the Picard rate, which does not seem to be the case here in most cases. Further, the Armijo method should work in a robust way for convex problems like this and it is strange to see such poor convergence. Moreover, very little importance seems to be assigned further exploring these non-expected results.

We added the convergence behavior of Newton's method with Armijo step sizes for different resolutions (Fig. 6). For higher resolutions, the relative difference decreases more. We also added the Picard iteration with Armijo step sizes. This algorithm stalls at nearly the same relative difference as Newton's method with Armijo step sizes. Therefore, the minimum of the convex functional and the solution of the full-Stokes equations seems to be too different for this resolution. The approximation of exact step sizes seem to have less difficulties with lower resolutions.

The non-quadratic convergence rate originates from small δ values, see the Figure below. [Hirn2013] discussed accuracy problems for small δ values. However, we think using small δ values is more suitable for ice models as those also use small δ values and the best choice of δ is also problem dependent regarding to [Hirn2013].

The Picard iteration would stall at the relative difference 0 at iteration 80 as the reference solution is the Picard iteration with 80 iterations.

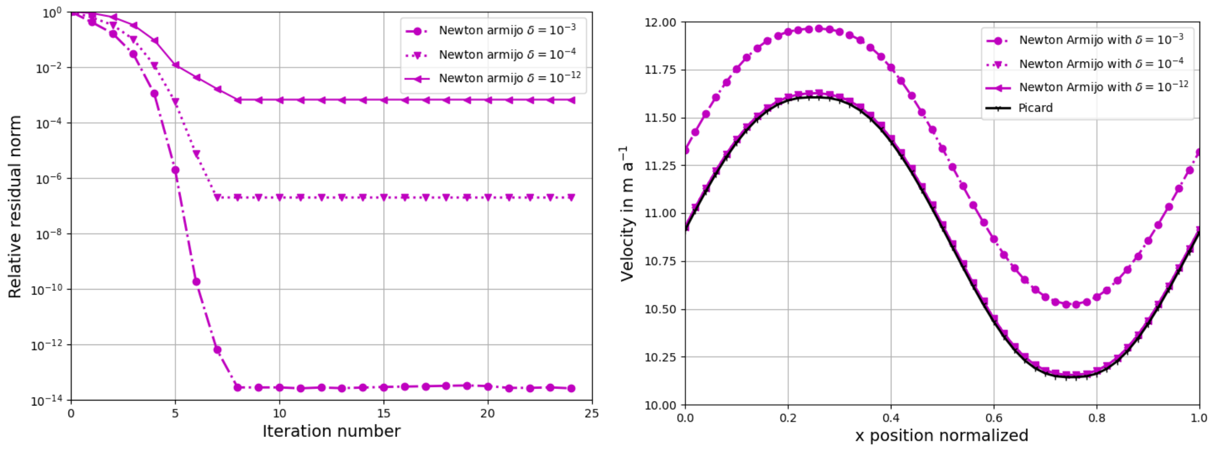

Figure 1: Left: Relative residual norm for different δ values. Right: Surface velocity for different δ values. The resolution for solving the experiment is given by 351 grid points in the x-direction and 10 grid points in the y-direction.

Based on this situation, and given the review from the first referee, significant improvement needs to be done to the current manuscript in order to be receivable in GMD. I would suggest, besides following all suggestions and addressing all critics from the first reviewer, you implement the following modifications:

- Extend the introduction to add context, comparison with existing solution strategies in other ice flow models, discuss limitations and better place your work in the glaciological modelling framework.

We added a summary of solution strategies in other ice models using the full-Stokes equations (lines 23-31).

- Check your code implementation as the results you report look suspicious. There is still a bug that prevents quadratic convergence of Newton's solver.

We made some further experiments to discuss the convergence of Newton's method. We changed the resolution: With a higher resolution Newton's method with Armijo step sizes reduces the error more. We think this behavior occurs as we have $J' = G$ in the continuous setting. However, in the discrete setting, we can have $J_h' \neq G_h$. Additionally, the Picard iteration with Armijo step sizes has the same convergence problems as Newton's method with Armijo step sizes (Fig. 4). Newton's method converges faster than the Picard iteration. It does not converge quadratic. We think that this originates from too high values of δ, as [Hirn2013] observed accuracy problems for small δ values and we added an experiment in [Schmidt2023] with higher δ values which resulted in faster convergence speed. We can not display the latest version, we submitted to this journal. Instead, we present quite similar figures to the submitted ones in this response, see Fig. 1. On the left plot, we see the relative residual norm for different values of δ. We see that the convergence rate is quadratic for a larger value of δ. On the right plot, we see that for $\delta = 10^{-4}$ the surface velocity is similar to the surface velocity for $\delta = 10^{-12}$. Thus, a larger value of δ can produce similar results but leads to faster convergence with Newton's method.

- Provide further details about the implementation and performance of the solvers. The convergence plots are for sure interesting, but are not the only results to report from your study.

We added the computation time for each iteration and the step size control for the three-dimensional experiment and the time-dependent experiments (Table 1, 3, and 5).

- Better motivate your various choices at all stages in the manuscript, providing some additional and relevant references in some fields.

We added a motivation for the choice of γ in Algorithm 3 (lines 127-128), and for a, b, and the number of the steps in the for loop in Algorithm 4 (lines 140-142). We also added a motivation for the initial guess (lines 165-167), and the choice of c for the relative local difference (lines 193-194).

- Provide a much more "in-depth" analysis of your results. If no change is observed after checking the code, it may be interesting to compare the behaviour of the solvers on other traditional benchmarks, such as viscous inclusion setup or others.

We think that we explained the behavior with different resolutions, refering to [Hirn2013]. We added the ISMIP-HOM experiments $E1$ and $E2$ with and without time-dependence.

Finally, it would be valuable to know your position regarding the preprint from 2023 which is very similar to this paper and may have already been submitted to a more math-oriented journal.

A new version of the preprint on ArXiv is in review at a more math-oriented journal. However, this manuscript has the following differences:

1. We considered additionally the experiments ISMIP-HOM A, $E1$, $E2$, and time-dependent versions of $E1$ and $E2$.
2. The experiments A and B have a local mesh-refinement to reduce the computation time for experiment A and experiment B has the refinement to make it comparable to experiment A.
3. We compared the relative difference and the local relative difference for the experiments A and B. In the mathematical manuscript, we used the residual norm as our error estimate.
4. This manuscript states the applied algorithms in more details with less mathematical termini. Moreover, the algorithms in the mathematical manuscript are formulated for divergence-free elements. Thus, all pressure terms vanish. In this manuscript, we state all this pressure terms.
5. We added the term $-\int_\Omega p \operatorname{div} \boldsymbol{v} \, dx$ to the functional as this is necessary for time-dependent simulations.
6. The target group for this journal is different to the mathematical manuscript. We aimed at making the key ideas and algorithms better understandable and reusable. Thus, we highlighted the used algorithms.

Taking the time and making the effort to carefully revisit and substantially extend the current work may provide a valuable input for the geoscientific modelling community and could be suited for GMD. However, in the current state, the work seems closer to a rushed submission than a complete paper.

We are convinced that our revision substantially increased the quality of the manuscript and are interested to hear if we could resolve all initial concerns. We are happy to discuss resulting issues and are hopeful to produce a manuscript that is suitable to GMD.