

Dear Ludovic Räss,

Thanks for the help in finding reviewers, reviewing, and extending the submission deadline.

Below are the two review responses from the discussion. We hope these responses, the reviewed manuscript, and the tracked changes answer all your initial concerns. If there are still issues left, we are happy to discuss them.

Review 1

We thank you a lot for the helpful comments and ideas. As suggested, we added the glacier d’Arolla from the ISMIP-HOM experiments in a time-dependent formulation. We also summarized time stepping methods for full-Stokes models. We modified the title to reflect the good performance of the Picard iteration with the approximately exact step sizes.

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

Summary: The paper implements a Newton method with step-size control for solving a full Stokes model describing the dynamics of glacial ice. In idealised diagnostic numerical experiments two types of step size control are compared with each other and to Picard iterations with and without step-size control.

General comments:

The chosen topic is important and the paper is well structured and seemingly without any obvious technical mistakes. My main concern is about the novelty of this paper, for the following reasons:

1. The Newton method is faster than Picard is expected and Newtons method is already used in ice sheet models extensively, even with step size control (e.g. in Elmer/Ice). It is true that Newton iterations do not converge for all cases especially without running a few initial Picard iterations, but for the numerical experiments chosen in this paper my guess is that Newton works in most existing codes already. Perhaps exact step size is new but I think in that case a brief summary of the most common step size control methods in the most used ice sheet models should be provided. Furthermore, a significant superiority of exact step-sizes compared to Armijo is not clear from this paper.

We added a summary of the step size controls and methods (Newton/Picard) used in full-Stokes models (lines 23-31). As recommended later, we added time-dependent experiments (pages 14-23).

2. In comparison to the previous manuscript which is referred to in this paper: Schmidt, 2013 <https://arxiv.org/pdf/2307.02930.pdf> In Schmidt, 2013 more mathematical details and proves are given, but the method and experiments seems identical. Is Schmidt, 2013 indented to be published?

We intend to publish this manuscript. In fact, a new version with some corrections is in the second review round at *Numerical Algorithms*. A preprint is on *ResearchSquare* at <https://www.researchsquare.com/article/rs-3354498/v1> . Due to your feedback, we added time-dependent experiments (pages 14-23) and references to glaciology models (lines 23-31) to this manuscript. These changes make this manuscript more different to the mathematical manuscript.

Nevertheless, I think the paper can be published after major revisions. My suggestions for addressing my above concerns are:

1. If Schmidt, 2013 <https://arxiv.org/pdf/2307.02930.pdf> is not already submitted somewhere, I think the mathematical proofs here are novel enough for publication. I am not sure if the

Cryosphere is the best journal for this, but possibly the mathematical proofs could be explained and summarized in the main text, and the proofs themselves be kept in the appendix.

2. Otherwise, a solution which probably fits the Cryosphere better, is to extend the numerical experiments and literature study so that the importance to glaciology is clearer. If the authors can show that it is possible to use Newton's method in realistic simulations in an effective way without initial Picard iterations, I think it is a major contribution to the community.

We follow this approach.

Specifically the following points should be included:

- (a) In the introduction, go through the most important ice sheet models (especially those using the full Stokes model) and review what method they are using to resolve the non-linearity of the problem

We summarized in the introduction, how *ElmerIce*, *ISSM*, *COMSOL multiphysics*, and *FELIX-S* solve the full-Stokes equations (lines 23-31).

- (b) Extend the numerical experiments. If the aim of this paper is to show that some method is better than another and could be useful for an ice sheet modeller, it must be for a relevant example. Most importantly, a time-dependancy (free surface movement) should be introduced. This is important since the first time-step (the only one studied here) usually behaves quite differently in terms of number of iterations, as compared to later time-steps when free surface has relaxed. In a real application, there is also a contact problem between the ice and bedrock, which introduces an extra non-linearity. The more realistic example (3D, variable bedrock) the better. Also, I think CPU-times should be measured rather than only number of iterations.

We added the two-dimensional glacier d'Arolla with variable surface and also with a sliding-part as in the ISMIP-HOM experiments *E1* and *E2*. We measured the computation times for the time-dependent experiments and the three-dimensional experiment.

If the original paper Schmidt, 2013 <https://arxiv.org/pdf/2307.02930.pdf> is kept as a separate paper, I think still some of the mathematical details must be added to this paper. I am lacking some mathematical justification of some of the key points of the paper, for instance the justification of the extra diffusion term in the equations, and explanation behind Algorithm 3.

We rearranged the section about step size controls (pages 5-6). We introduced the Armijo step size (lines 122-129) first. Then, we motivate the exact step sizes (lines 130-142). Additionally, we explained the algorithm for the exact step sizes a bit more (lines 132-134).

We stated the reasons for adding the diffusion term (lines 81-82) and explained this in the appendix (A1 The variational formulation).

Specific comments:

The abstract and introduction could flow better. The sentences could be longer or better connected.

We changed the abstract due to additional content and tried to connect the sentences better. We also changed the introduction a bit to get a better flow.

L 4 and 5: I think these sentences can be removed from the abstract "For the step size control, we need a minimization problem. Minimizing a specific convex function is equivalent to solving the full-Stokes equations."

We removed these sentences.

L 18: nonlinear \rightarrow non-linearly?

We corrected the grammar.

L 18: make it clearer that the stationary variation of these equations is the standard one

We added, how the new shape of glaciers is calculated. This should make clearer why the stationary equations are important (lines 18-20 and in detail in section 6.2).

L 56: These spaces are for linear Stokes, see e.g. the work by Belenki et al:
<https://www.jstor.org/stable/41582741> for the appropriate spaces

Thanks for the reference. We added an explanation in the appendix (A1 The variational formulation) with reference to this literature (lines 359-360) and explain that the additional diffusion term μ_0 allows us to use the simpler spaces for linear Stokes.

L 60: Why do you add a diffusion term? This is non-standard and not included in other ice sheet models, which might make the reader wonder if these results are applicable to their model. I suggest trying to get rid of this term.

We added a short explanation for the diffusion term in the appendix (A1 The variational formulation). Additionally, we did all experiments again with $\mu_0 = 0$ (and added this information in line 85).

However, we partly overwrote our reference solution by our methods that we compared. This led to small mistakes in Fig. 4 and 7 for Newton's method and the Picard iteration with exact step sizes. Now, both methods are a bit better for the later iterations. Additionally, Fig. 3 was slightly wrong for the exact step sizes. Now, the Picard iteration with exact step sizes has a slightly higher difference at the surface to the reference solution and Newton's method with exact step sizes a slightly lower.

Regarding the diffusion term in theory: Without the diffusion term, we have to formulate the problem in a subspace of $W^{1,1+1/n}(\Omega)^N$. For functions $f \in W^{1,1+1/n}(\Omega)^N$, we know

$$\int_{\Omega} |f|^{1+1/n} dx < \infty \quad \text{and} \quad \int_{\Omega} |\nabla f|^{1+1/n} dx < \infty.$$

This is not enough for the well-posedness of the first summand in the Picard iteration and the directional derivative of the variational formulation Eq. (6). The additional diffusion term allows us to have $f \in H^1(\Omega)^N = W^{1,2}(\Omega)^N$. Then, we know

$$\int_{\Omega} |f|^2 dx < \infty \quad \text{and} \quad \int_{\Omega} |\nabla f|^2 dx < \infty.$$

Now, we can apply Hölder's inequality to show $\int_{\Omega} |Dv_{k+1} : \nabla \phi| < \infty$.

L 71: I would not say that Picard is standard. Include some references for ice sheet models that use Picard rather than Newton. Most probably use a combination of the two.

We added ice models, which use the Picard iteration and replaced "standard" by "common".

L 92-93: It's true that in Hirn μ_0 is 0 (and even $\delta=0$!) . This should be discussed - why is it not possible to set them to zero in this work?

We changed the simulations to $\mu_0 = 0$. Only for the infinite-dimensional theory $\mu_0 > 0$ is necessary. For the first few iterations, the solution is nearly identical. Some small changes are visible for more iterations.

Algorithm 3: what is the stopping criteria, i.e. how many for-loop iterations do you do?

We used 25 iterations and added this in the text (lines 140-142). One could probably optimize the number of used iterations. However, in terms of the computation time, this is not important.

L 104: For the reader of the cryosphere, I think you should introduce the two step-size control methods with a bit more words, discuss pros and cons etc, rather than just stating the algorithm.

We motivated Armijo and exact step sizes more. Additionally, we introduced the Armijo step sizes first (lines 122-129) and used them as a motivation for the exact step sizes (lines 130-142).

Equation 10-11: Does the fact that the terms are smaller really mean that one can surely say that the solution is not impacted significantly, especially in a time-dependent simulation it is not clear to me.

We set $\mu_0 = 0$ for the simulations. The term $\delta > 0$ is used in ice models. Thus, we did not set $\delta = 0$. However, as we have for typical ranges of $|D\mathbf{v}|$ in the computer the relation $D\mathbf{v} = D\mathbf{v} + \delta$ due to the machine precision, the δ term should not influence this region. It will have an influence on really small values of $|D\mathbf{v}|$, but there are also other inaccuracies due to, e.g., grid resolution, unknown Ω , B , or N . Thus, we think that a small value of δ does not change the solution too much compared to the other sources of inaccuracies.

Figure 2: I think it is misrepresentative to compare to models that are not full stokes, please only include full stokes models for the dashed lines.

We specified in section 5.3 that we only compared with full-Stokes models. (*aas2*, *cmal*, *jvj1*, *mmr1*, *rhi3*, and *ssu1*). The displayed standard deviation in Fig. 2 seems to us quite similar to Fig. 6 in [Pattyn2008], just with another scaling on the y -axis. However, we forgot the models *aas1*, *oga1*, and *rhi1* and added them calculating for the dashed lines.

About the experiments: I strongly recommend that you add a more complex example, something with more complex geometry and time-dependency. A time-dependent simulation of the arolla glacier maybe.

We added the experiments ISMIP-HOM *E1* and *E2* in a time-dependent formulation. For the time-dependent simulation, we had to add

$$- \int_{\Omega} p \operatorname{div} \mathbf{v} \, dx$$

to the functional in Eq. (7) because for the time-dependent problem our initial guesses are not necessary divergence-free.

About the experiments: Add plots with relative error vs CPU-time, to show if the step-size control increases the work per iteration significantly or not

We did not do a relative error vs computation time plot for two main reasons: The computation times for all methods are quite similar for the time-dependent experiments (Table 3 and Table 5). For the three-dimensional example, Newton's method itself takes more time than the Picard iteration (Table 1). However, the step size control is not computationally relevant. Thus, we think there would be too many details in the plots without giving important information. Instead, we added the tables with computation times to give this information. Secondly, one could reduce the computation of the step size control by parallelizing the calculation of the integrals. The *MUMPS* solver automatically uses more processors, if available. Thus, we have a comparison between one optimized algorithm and one proof of concept algorithm. Additionally, one could try if a lower number of bisections for the exact step sizes produces results of similar quality. Therefore, the computational effort for calculating the step sizes could be reduced compared to the complete iteration.

The outlook: This is a little bit informally written.

We rewrote the outlook by adding some possible research directions and reformulating the text a bit more formally.

Additionally, we made minor corrections due to mistakes in the reference in Schmidt, 2023: We

removed the statement that the function G would be continuously differentiable and the theoretical statements about the convergence speed. The function G is only Gâteaux differentiable.

Review 2

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.

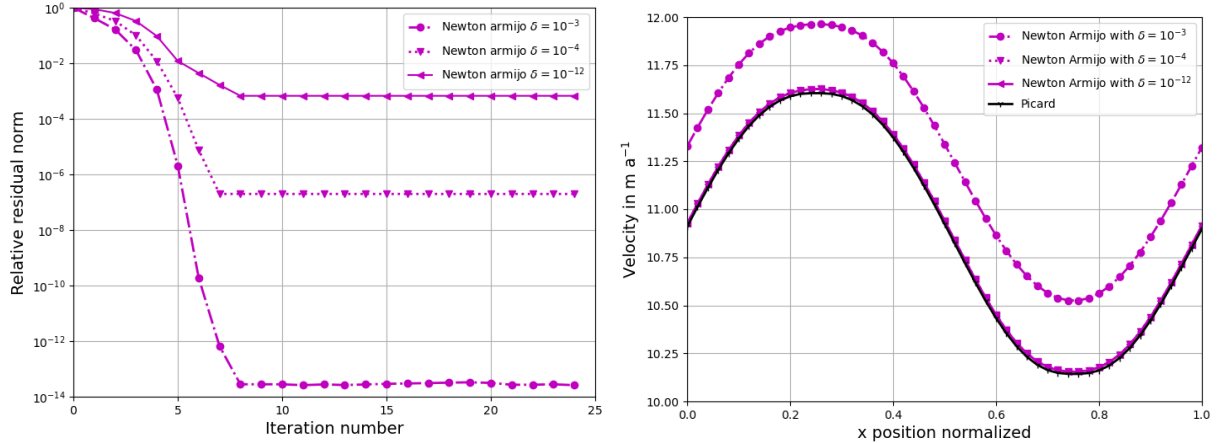


Figure 1: Left: Relative residual norm for different δ values. Right: Surface velocity for different δ values. The resolution for solving the experiment is given by 351 grid points in the x-direction and 10 grid points in the y-direction.

[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.

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, referring 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.