Nathan Martin
Toulouse Institute of Mathematics - UMR CNRS 5219
Dept. GMM, INSA, 135 avenue de Rangueil, 31077 Toulouse Cedex 4, France
nmartin@insa-toulouse.fr

Dear Dan Goldberg,

Please find hereafter the answers to your comments, concerns and remarks regarding the article entitled :

"Of the gradient accuracy in Full-Stokes ice flow model: basal slipperiness inference"

by Nathan Martin and Jérôme Monnier

Sincerely yours,

N. Martin and J. Monnier

You expressed an important concern about the level of english in the paper, as the other referees did. In order to address this concern, a thorough attention to the correct and sensible use of english will be paid in the revised version of the paper. However, although the following comments have been carefully written, they might contain english mistakes and we apologize in advance.

1 Broad comments on methodology and content

- The use of a spectrum with localized wavelength is useful to assess with a certain level of precision the level of basal variability one can retrieve from surface velocity observations. Therefore the use of a discrete spectrum is sufficient to estimate orders of magnitude and provide lower-bounds.

  It seems natural that the use of satellite-inferred data lessen the ability to distinguish the behavior of both methods since, as you already mentioned it, the spectrum is probably more continuous and complicated than the manufactured one we consider in this work.

  However, the fact that a small difference is observed between the exact and approximate adjoint in terms of lowering j (*i.e.* the final cost of the identification process) is an insufficient information to compare the results of both methods (typically in Morlighem et al. 2013). The lines 9-13 p.3868 of the present work tried to recall this aspect. However, as you mentioned it later, this paragraph seems unclear. So here is a possible reformulation:
  "In all cases, the final cost reached by both methods is not a sufficient information in order to distinguish their precision, especially for noise level greater or equal to 1%. On the contrary the frequency analysis allows to detect that an identical final cost is not equivalent to an identical inferred friction coefficient. It demonstrates the ill-posedness of this type of inverse problem which can be seen as an equifinality issue (*i.e.* the fact that an identical state, and consequently an identical cost, can be obtained with different sets of input parameters)."

  Similarly, a simple relative error (typically $|\beta_f - \beta_s|/|\beta_f|$) between the exact adjoint β_f and the "self-adjoint" one β_s is unable to show the differences between both methods. That is why we pay little attention to those quantities in the present work in order to focus on the frequency analysis that provides a much clearer way to compare the results of both methods.

  The present work tries to establish lower bounds on the wavelengths an adjoint-based inverse method is able to retrieve from surface velocity observations. As we can see, these lower bounds depends on several parameters such as the noise, the density and the slip ratio. These results can therefore help to know the limits of the approach and then help to distinguish, in an identified friction coefficient, what can be considered as a real component of the friction at the bedrock (and thus interpreted) and what cannot be considered as an actual phenomena because of the numerical limitations of the method.

  However the data does remain synthetic even if the topography of the bedrock and the shape of the surface are real field data.

  As you recommand it, Morlighem et al. 2013 should be cited (it was not published when the present paper was submitted) in the introduction after the reference to Goldberg and Sergienko 2011 as a work comparing the exact and self-adjoint method for the higher-order model. A comment on their results, compared to ours, could be added to highlight what is developed herebefore. For instance, after the paragraph p.3868 l9-13, the following sentence could be added: "In Morlighem et al. JGR 2013, the authors observed that the use of the "self-adjoint" approximation barely affects the convergence of the inversion process and thus makes of the "self-adjoint" approximation an excellent approximation of the exact adjoint. This observation does not take into account the equifinality aspect of the inversion procedure. It is important to point out that the poorer the data (or similarly the greater the noise), the stronger the equifinality."

- A more technical paper describing a new algorithm for the solution of the power-law Stokes problem, written by the same authors, now in minor revision, should be cited here to address the details about DassFlow. However some more detais could easily be given in an appendix, for instance about the way the adjoint of the linear solve is performed. It is a bypass approach, very similar to the one detailed in Goldberg and Heimbach TCD 2013. In the meantime the details can be found in the first author PhD thesis that appears in the references (that can be found on internet, http://www.theses.fr/en/2013ISAT0021). A part of the manuscript is in french, however the details about the bypassing of the linear solver are given in appendix A in english. The linear solver used is MUMPS. An archive version of the other paper can also be provided.

  On the other hand, some rough estimations of the computation time for the forward and backward solve could be given as an indication *e.g.* of the time saving one can expect.

- The required truncation to maintain a maximum accuracy strongly depends on the situation (in terms of data precision) of each numerical experiment. The truncation of the forward run made in the gradient test section, in order to assess the precision of the adjoint, observe the rate of convergence of the reverse accumulation and the robustness of the the truncated adjoint, is rather numerical and should not be considered as a reliable way of saving computation time. However, the point of the truncated adjoint approach is to propose a method to adjust the accuracy of the adjoint based on the setup of the numerical experiment. Although it is difficult to deduce a systematic and quantified way of truncating the adjoint with respect to the noise, the advice that we would propose for real numerical simulations is to perform an accurate direct simulation, regardless of possibly noisy data but an adjoint in adequation with the noise. This adjustment can be made based on one gradient test which allows to quantify the rate of convergence of the reverse accumulation loop. By contrast, the slip ratio has a different influence on the identification procedure since it modifies, for a given level of noise, the shortest wavelength the adjoint method can retrieve. Consequently, the adjoint precision should not be adjusted based on this aspect.

2 Specific comments

- eqs (4-5): it will be precised in the revised version
- 3858 l9: it was suppose to mean "in what follows". It will be modified in the revised version.
- 3859 l2: As pointed out by Referee # Stephen Cornford, the sentence is not correct and should be reformulated. There is no implication between the fact that the cost function becomes more regular thanks to the regularization term and the fact that it introduces a bias toward smoothly varying β. The restriction to smooth β is due to the chosen regularization operator (which is the gradient). In addition, the fact that the cost function would be more convex is not necessarily true when using this regularization term. It would have been true if the regularization operator was the identity (and thus the regularization term the square of the norm of β). So in the present case the sentence could be replaced for example by:

"This term regularizes the functional to be minimized and introduces a bias toward smoothly varying fields".

- 3862 l1: We could simply say "In order to validate the adjoint code..."
- 3862 l5: It will be corrected in the revised version
- 3863 l13: idem
- 3863 l26: idem
- 3864 l7: It was suppose to mean "the use of such a method is equivalent to retain"
- 3864: It will be done in the revised version. It should also be plotted as an illustration
- 3865 l26: It will be corrected
- 3867 l3: idem
- 3867 l6: The use of "a priori" is probably more common in french. Here, it was suppose to mean "The number of unnecessary iterations is likely to be strongly dependent on the situation..."
- Section 3 and figs 2-3: In equation (17), $\delta \boldsymbol{k}$ is the perturbation direction or, similarly, the direction of derivation (whereas α represents the amplitude of the perturbation). The notation is maybe a bit misleading because in the Subsection 2.4 describing the derivation of the adjoint model, we introduce $d\boldsymbol{k}$ which is in fact equal to $\alpha\delta\boldsymbol{k}$. The gradient test then makes α vanishing (the amplitude) while maintaining the direction $\delta\boldsymbol{k}$. It should be precised in the revised version.
- 3868 2$^{\text{nd}}$ paragraph: hopefully, the details given in the present letter in the first section (precisely those about this paragraph) made it clearer.
- 3868 l4: "sloppy" was supposed to be "sloping" or inclined. It could be more rigorous to call it a channel flowline for example. It can be modified in the revised version.
- Eq (19): another referee also pointed out the confusion so it could be replaced by N for instance and precised that it is just a natural integer between 1 and 3. We are also thinking about a more compact and efficient way to write these lists of frequencies (here and in sections 4.2 and 5)
- 3869: it will be modified
- 3873 l4: it will be corrected and the revised version of the manuscript will try to provide an improved level of english.
- 3875: paragraph at line 21: discrepancy means misfit and what we call the "optimal discrepancy" is the smallest cost the Morozov's principle lead to choose depending on the regularization parameter ; a smaller discrepancy would result in over-fitting.
- 3876: it will be corrected
- 3876 paragraph at line 20: The results presented in figure 9 led us to conclude that the chosen frequencies were maybe too high for a 1% noise. For instance, the $2h$ wavelength cannot be considered to be well captured for every slip ratio, even with the exact method. So this sentence was trying to assess reliable lower bounds on the accessible wavelengths in this situation. Therefore, without actually plotting the result, we observed that a faithful reconstruction, for any slip ratio, can be obtained for wavelength of $10h$ and $5h$. However the sentence "shorter wavelengths are not accessible" is wrong as we can see on Figure 9. The wavelength $2h$ is accessible but not well captured. The paragraph should then be reformulated.

- 3877 l14: Do you mean in the plot? Since $r = 6$ shows a slightly higher gradient than $r = 13$ it seems coherent with the decreasing of the norm of the gradient with increasing slip ratio.

- 3877 l16: It would be indeed appropriate to show the topography when making such a statement. They will be plotted in the revised version. In the meantime, find on Figure 1 a possible plot of this geometry:

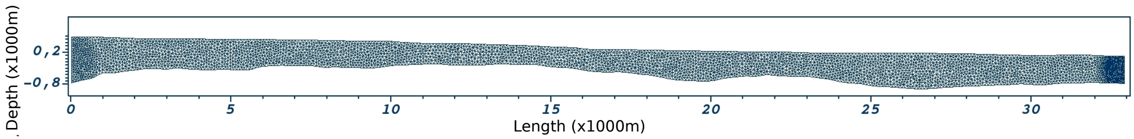

Figure 1: Vertical cut of the outlet glacier Mertz, Antarctica (topography profile from ICECAP 2010 within Ice bridge, provided by B. Legrésy, LEGOS, France) , x-scale $= 2/5$

- We mean the last 2 states *of the direct run*, or equivalently the first two states of the reverse accumulation loop. In other words, retain 2 iterations instead of one (in the case of the "self-adjoint" approximation) improves the precision of the "self-adjoint" approximation while maintaining a small computation cost.