Geosci. Model Dev. Discuss., 5, C1166–C1170, 2013
www.geosci-model-dev-discuss.net/5/C1166/2013/

Geoscientific Model Development Discussions

Interactive comment on "Technical Note: Improving computational efficiency in large linear inverse problems: an example from carbon dioxide flux estimation" *by* V. Yadav and A. M. Michalak

Anonymous Referee #2

Received and published: 4 January 2013

1 Overview

The manuscript by Yadav and Michalak presents algorithms for calculating matrix products that commonly arise in geophysical inverse problems. The subject matter is suitable for GMD, and the theoretical results and sample case appear to be of value in terms of significant computational advances over direct method. I do however have two main concerns. First, the presentation of the results could be better framed with regards to the computational literature. Second, based on tests using the sample code provided, the efficiency of the new algorithms over direct methods appears to occur under a more limited range of conditions that indicated by the theoretical analysis in the

manuscript. Below I expand upon these points, as well as a few other minor comments, that would be best addressed prior to publication.

2 General comments

- I appreciate the authors have provide MATLAB code to support their work. I tried to replicate the settings used for the case study based on my understanding of the variables as presented in the manuscript. For example, the manuscript states that $rt = m_s$, and that $m_s = 2635$ for the case study. Similarly, $pq = m_t$. Assuming a square system, using $r = t = p = q = 52$ would seem to be a reasonable approximation of their test case described in section 1.1. I then tested this with $n = 8503$ for the `HQ` test in `HQ_HQHt.m`. Repeating the experiment several times showed that the indirect method was slightly faster on average, by one second out of ten. How is this consistent with the claims of several orders of magnitude increase in efficiency in the manuscript? Granted my tests conflate floating point efficiency with memory efficiency, but I would have expected greater speedups for the indirect approach given the conclusions of the manuscript.

  Further, upon additional tests with the provided code, trying setups where $r = t$ and $p = q$, it seems that if m_s is smaller than m_t, then the indirect approach is less efficient, not more efficient, and vice versa. For example, if $r = t = 100$ and $p = q = 10$, then the indirect method was on average $\sim$30% faster, but if $r = t = 10$ and $p = q = 100$, then the direct method was faster by a factor of three. These were all using $n = 8503$, although as mentioned in the manuscript, n does not impact the ratio of floating point calculations used in the direct relative to indirect approaches.

  Therefor, I encourage the authors to consider cases beyond their one test case, and to present the efficiency of their algorithm in terms of the magnitude of m_s

versus m_t. At present it appears that either their new approach is only of benefit when $m_s < m_t$, or there is something that is not clear in the manuscript or provided code.

- There isn't any background on literature regarding computational algorithms in the introduction. Instead, the introduction focuses exclusively on the literature related to applications. While the latter is of value given the GMD audience, the former should also be included. Therefore, discussion currently included on page 3334 of other algorithms such as Strassen's, Coppersmith-Winograd, and Saibaba and Kitanidis should be introduced and explained.

- The claim that the matrix calculations HQ is the first bottleneck of inverse problems (p3328, line 8) also does not seem to be substantiated. The bottleneck in many approaches is the calculation of H itself, hence the development of gradient based approaches (e.g., 4D-Var) that do not require explicit construction of the full transport model Jacobian. The applicability of the results in terms of improving the computational efficiency of large linear inverse problems is thus perhaps more limited than the title would imply. For example, are results relevant most for specific approaches such as geo-statistical, Kalman filtering or variational methods?

- Following on the previous comments, since it appears from the structure of the original manuscript that consideration of other methods in the literature was somewhat of an afterthought, it isn't evident that the authors have done a thorough literature review of the computational aspects of their problem (perhaps they have, it just does not come across in the paper). For example, a recent paper by Singh et al. (GMD, 2011) presents work on efficient representation of covariance matrices using Kronecker products, including common linear algebra manipulations such as matrix-vector products and matrix inverse. How is the work of Singh related to the approaches presented here? At the very least it seems worth men-

tioning, particularly given the concluding sentence of the manuscript regarding the need for advances in the numerics of other aspects of geophysical inverse problems. Another work to consider that used Krockecker product representation of covariance matrices is Meirink et al. (2008). Another reference that may be of interest is: Sun, Y., Li, B., and Genton, M. G. (2012), "Geostatistics for large datasets," in Advances And Challenges In Space-time Modelling Of Natural Events, Springer, Vol. 207, Chapter 3, 55-77.

3 Specific comments

- It is a bit lazy to reproduce verbatim the sentences from the introduction in the abstract

- 3327.5 I believe the 2nd edition of Aster is copyright 2013. Also, it is a bit odd to use this as a reference for hydraulics and remote sensing; as I recall most of the examples are taken from tomography and seismology.

- 3327.10 Check the year on the Ciais citation.

- 3328 It could be useful to more clearly explain what is meant by aggregation error; Meirink et al. (2008) would be a suitable reference.

- 3328.15 Not a big deal, but curious why here the penalty term in the objective function is described as a "prior s_p" whereas the previous works by Goudji emphasized that explicit prior fluxes were not used, and that this term was $X\beta$.

- 3329.23 Could be useful to mention here the coding language.

- 3333 It is a bit odd to present the efficiency for the specific case as a ratio (eq 13), but later for the generic case as two equations (14 and 15). Why not just say for

the latter that the indirect approach is a factor of $2n^{1/2}$ faster for large n?. Overall, considering how the efficiencies are presented in both the results and abstract, at the moment there are ratios, percentages, order of magnitude comparisons, ... the paper may befit from picking a single metric for comparison and sticking without throughout.

Interactive comment on Geosci. Model Dev. Discuss., 5, 3325, 2012.