Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-325-AC1, 2021

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on “Lossy Checkpoint Compression in Full Waveform Inversion” *by* Navjot Kukreja et al.

Navjot Kukreja et al.

nkukreja@imperial.ac.uk

Received and published: 2 February 2021

We thank the reviewer for a very considerate review and the insightful comments that have helped us improve the paper. We would be happy to make the following changes in the revised version.

"The authors suggest that the idea of using compressed checkpoints is novel, and allows a tradeoff of time and disk space."

While the reviewer has pointed out relevant prior work in this area, we are not aware of prior work applying lossy compression in combination with checkpointing in seismic imaging. The details of inverse problems vary greatly between different application domains. The tradeoff we discuss is between peak DRAM footprint, total program

Printer-friendly version

Discussion paper

runtime, and absolute error tolerance. We do not discuss checkpointing to disk in this paper - the commercial seismic imaging computers are generally designed to have large memory and problems are configured to avoid any checkpointing to disk because the overhead is judged too costly. This is another example of how inverse problems are handled in different sectors.

"However, I don't see evidence of exploration of the runtime speeds that could be achieved, and under- standing of the accuracy/performance tradeoff beyond the schematic image in Figure 3."

We previously studied these aspects in previous work [1], which has been duly cited in the current work. During the previous work, which is generally applicable to adjoint-based optimisation problems, we saw that it is possible to accelerate adjoint computations, depending on the compression factor in the lossy compression. For lossy compression, the compression factor depends on the acceptable error tolerance, and the error tolerance is highly specific to the problem being solved. This current paper is an empirical study of the effect of lossy compression, specific to FWI - i.e. evaluating the impact of lossy compression of the stored wavefield on the accuracy of the computed gradient and overall convergence of the inverse problem.

"Further, I would expect that this type of approach to be explored elsewhere and be a more minor contribution, particularly since in industry full storage of the forward solution appears to be the norm in FWI."

The reviewer is making a reasonable assumption - this may well have been explored in commercial seismic imaging. However, there does not appear to be anything published in the open literature evaluating the impact of lossy compression on seismic inversion. Indeed, a common crude compression technique used in industry and academia is to subsample in time. We have not found any quantitative analysis of this approach in the literature, which is surprising when you see the high level of error incurred using that strategy (see figure 24, section 4.7). The closest work that we are aware of

Printer-friendly version

Discussion paper

that performs similar analysis is [2] which performs the cross-correlation (during back-propagation) in Fourier space. That method is implicitly compressive because it only stores a small set of wave numbers. This method has been known about for some time but not frequently discussed quantitatively in literature - this is possibly due to patent CA2760053C on the use of the method in tomographic image acquisition and reconstruction.

"I was hoping that the work would be more generally applicable to PDE-constrained optimization, but due to the use of a linear PDE constraint (which is consistent with FWI) I have my doubts about the extensibility of the results to cases involving turbulent flow, for instance. This is due to the need to linearize the PDE around a state to compute the gradient."

Inverse problems are hard - particularly when dealing with the realities of real data, noise and inexact physics. To optimise the computation of inverse problems (total compute resources and/or time to solution) we have to exploit domain specific knowledge and there is an element of engineering a practical solution. We know that for related problems such as data assimilation in weather/climate science, small perturbations can have a detrimental impact on the solution - in some cases bitwise reproducibility is arguably required. For this reason, we have made it clear in the title and throughout the paper that we are only investigating the use of lossy compression in the domain of seismic imaging which we know to be less sensitive to perturbations compared to chaotic dynamical systems for example.

"The work Weiser and Götschel (See State trajectory compression for optimal control with parabolic PDEs, SISC, 2012) appears to be as complete an empirical study for parabolic systems, with an additional error analysis. "

The paper the reviewer has pointed us at (Weiser and Götschel) is indeed a very valuable addition to the literature review and is certainly relevant. A reference to this paper will be included in the next revision. Weiser and Götschel discusses general parabolic

PDEs, while the current work discusses only FWI based on the acoustic wave equation (hyperbolic), going into detail about the propagation of errors through intermediate steps. We also note with interest that Weiser and Götschel devise a new compression algorithm. We focused on use of ZFP as a general-purpose compressor for floating point data as it has the advantage that it's now widely ported. In Section 4.2, Weiser and Götschel talk about compression as an alternative approach to lossy compression. However, by looking at checkpointing and lossy compression in combination, our approach opens up an entire spectrum between these two choices using a tunable cost model.

"I'm not sure how to reconcile this with the statement in the present work that the analysis of compression errors is "beyond known numerical analysis"." We agreed this was phrased poorly and we will correct this in the next version.

"In the review, you could cite the growing work in parallel-in-time optimization by for instance S. Gunther or S. Gotschel (going back to the early 2000's Mayday and later Ulbrich and Heinkenschloss)."

We are aware of the exciting research into exponential time integrators - particularly as applied to weather/climate modelling as a potential way to improve strong scaling while having the additional benefit of necessitating the storage of fewer snapshots during data assimilation. However, to the best of our knowledge, it is still an open research question as to how time varying source (and receiver) terms can be handled in such a formulation when the model time step is near the Nyquist rate required to preserve the signal contained in the (seismic) source terms. If a solution for this mismatch in time scales were resolved for seismic applications then we would see exponential time integration as an additional method to be used in conjunction with these other methods depending on the context. With these points in mind, we currently consider these schemes to be orthogonal to the focus on this paper.

"How is the signal to noise ratio measured?"

We use the following definition for Peak Signal to Noise Ratio (this will be included in the revised version): (Image attached) Where MAX_I is the range of values being compressed (maximum-minimum), and MSE is the mean squared error between the reference solution and the lossy field.

"Could there be some more discussion about the wave energy included? It seems that compression would be a dissipative mechanism in general (removing hopefully high frequency modes). How does this impact the objective? Is it systematically less if so why or why not?"

ZFP's impact on the frequency spectrum of compressed images was previously studied in [3] (see Fig 13, for example). We will add a suitable reference in the paper. It appears from these experiments that compression is actually a benefit. We speculate that lossy compression is acting as a form of regularisation - ie less significant bits are dropped and thereby smoothing effect to the field. As can be seen in [3] the reconstruction errors are normally distributed - so thinking about this as a dissipative mechanism is reasonable. Regularisation is a complicated topic and while we can discuss these observations in the conclusion we are reluctant to make any claims in this regard at this point in our research.

"The structure of the figures, with one caption per image as opposed to multiple (and some error plots relegated to the appendix) is frustrating to read. This lends itself to the overall feeling that this is a lab report on this material."

The figures are in that format as this is what we understand to be required by the GMD submission guidelines for latex - https://www.geoscientific-model-development.net/submission.html#templates "Please provide only one figure file for figures with several panels, and please do not use \subfloat or similar commands."

References [1] Kukreja,N., Hu ÌĹckelheim,J., Louboutin,M., Hovland,P. ,Gorman,G.:Combining checkpointing and data compression to accelerate adjoint-based optimization problems. In: European Conference on Parallel Processing. pp. 87–100.

Printer-friendly version

Discussion paper

Springer (2019) [2] Witte, P. A., Louboutin, M., Luporini, F., Gorman, G. J., and Herrmann, F. J.: Compressive least-squares migration with on-the-fly Fourier transforms, Geophysics, 84, 1–76, 2019 [3] Tao, D., Di, S., Guo, H., Chen, Z. and Cappello, F., 2019. Z-checker: A framework for assessing lossy compression of scientific data. The International Journal of High Performance Computing Applications, 33(2), pp.285-303.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-325, 2020.

$$
\begin{aligned}
PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\
&= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\
&= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE)
\end{aligned}
$$

Fig. 1.