High-performance backpropagation in Scientific Computing

Navjot Kukreja [1] Jan Hückelheim (advisor) [2] Paul H J Kelly (advisor) [1] Gerard J. Gorman (advisor) [1]

Imperial College London [1], Argonne National Laboratory [2]

Motivation - Seismic Imaging

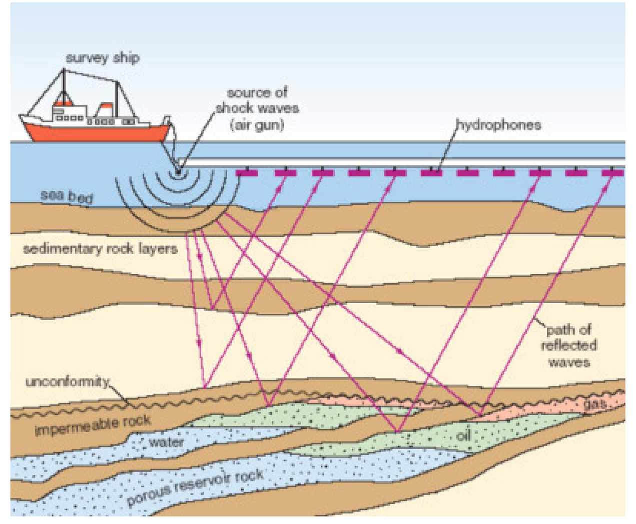


Figure 1: Offshore seismic survey (Source: OpenLearn)

The wave equation (Equation 2) can be solved to simulate the propagation of seismic waves through the earth. Given a source signal and the earth's physical parameters, we can solve the PDE using a numerical scheme (in this case explicit finite difference with Devito) to simulate the signal received at the hydrophones.

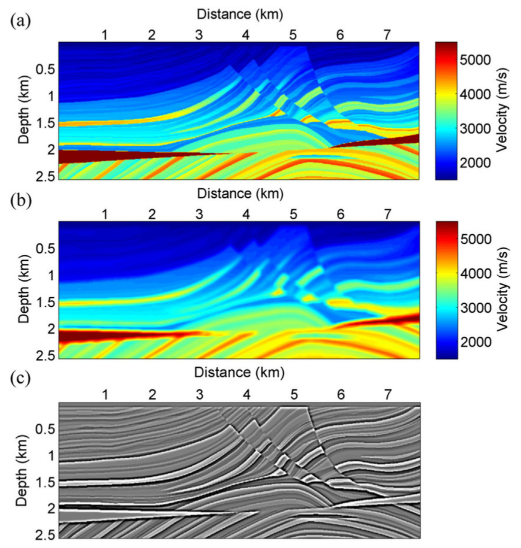


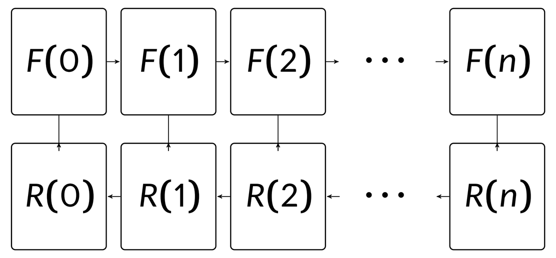
Figure 2: Marmousi velocity model

To convert from the received signal (Figure 1) to an image of the earth's subsurface (Figure 2), we can set up an optimization problem (Full Waveform Inversion) that finds the values for the earth's physical parameters that minimize the difference between the simulated signal at receivers and the observed signal.

The gradient of the least squares objective function with respect to the model parameter **m** is given by [1]:

$$\nabla\Phi_s(\mathbf{m}) = \sum_{t=1}^{n_t} \mathbf{u}[\mathbf{t}]\mathbf{v}_{tt}[\mathbf{t}] \qquad (1)$$

where $\mathbf{u}[\mathbf{t}]$ is the wavefield in the forward problem and $\mathbf{v}_{tt}[\mathbf{t}]$ is the second-derivative of the adjoint field. The computation of this gradient is what we call backpropagation.



We would like to address some performance issues inherent to this backpropagation, in a Domain-Specific Language, automatically.

The Devito Domain-Specific Language

Devito is a Domain-specific language for the automatic generation of high-performance finite-difference solvers.

$$\begin{cases} m\frac{d^2u(x,t)}{dt^2} - \nabla^2 u(x,t) = q_s \\ u(.,0) = 0 \\ \frac{du(x,t)}{dt}\big|_{t=0} = 0 \end{cases} \qquad (2)$$

```
e = m * u.dt2 - u.laplace
stencil = Eq(u.forward, solve(e, u.forward))
op = Operator([stencil], ...)
op.apply(...)
```

```
void finite_difference_solver(...) {
  //...impenetrable "performance optimized" code
}
```

Automatic Differentiation of Stencils

Problem 1: Gradient must be formulated by hand - restricting the possible choices of equations and objective functions. Automatically generated derivatives of stencils (gathers) are scatter operations.

Solution 1: Automatically generate (parallel) derivatives for stencil operations and rearrange them back into stencils.

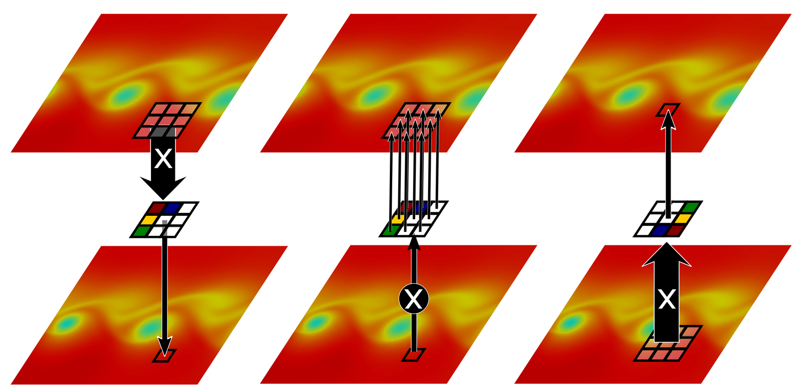


Figure 3: Parallel derivatives of stencils

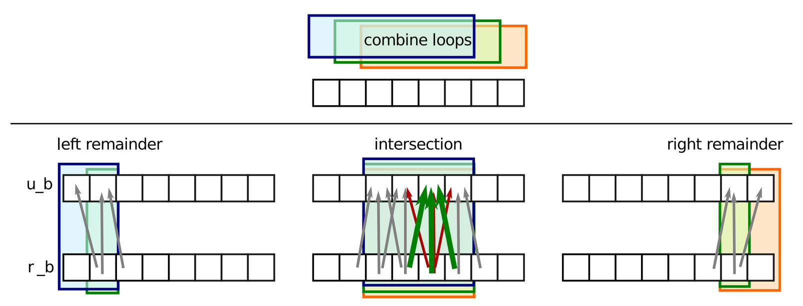


Figure 4: The transformation to convert the scatter back into a gather

Checkpointing

Problem 2: Storing intermediate states from the forward computation requires too much memory.

Solution 2a: Save a subset of states and rerun the forward computation when required. (Memory-compute tradeoff)

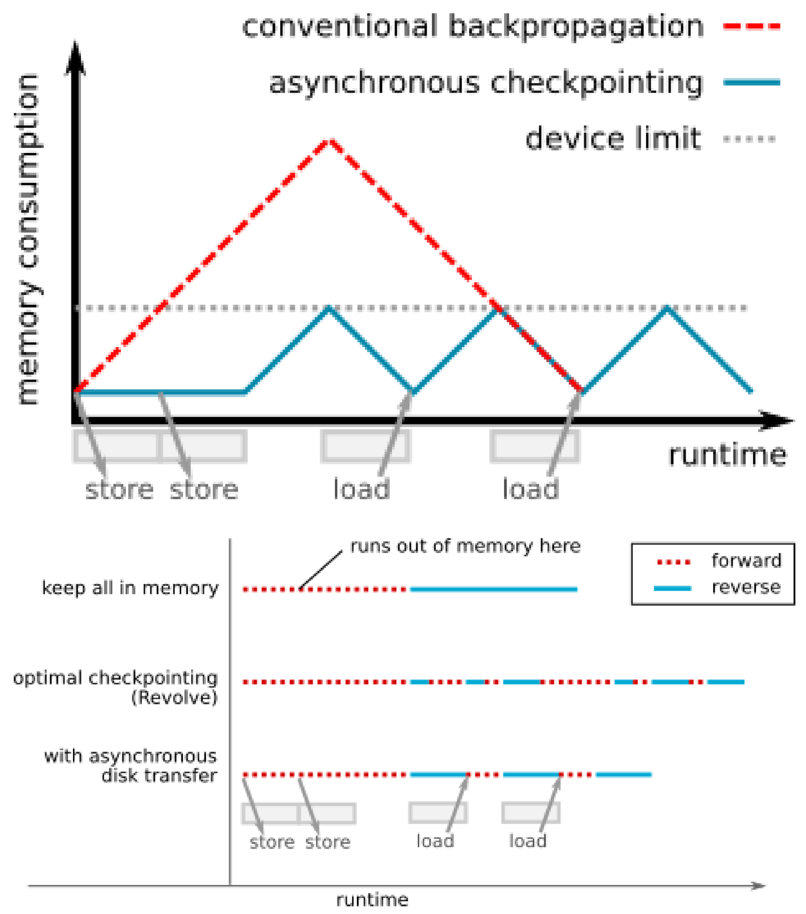


Figure 5: Timeline for conventional adjoint, Revolve checkpointing, and asynchronous multistage checkpointing.

Lossless/Lossy Compression

Solution 2b: Combine checkpoint-restart with lossless/lossy compression.

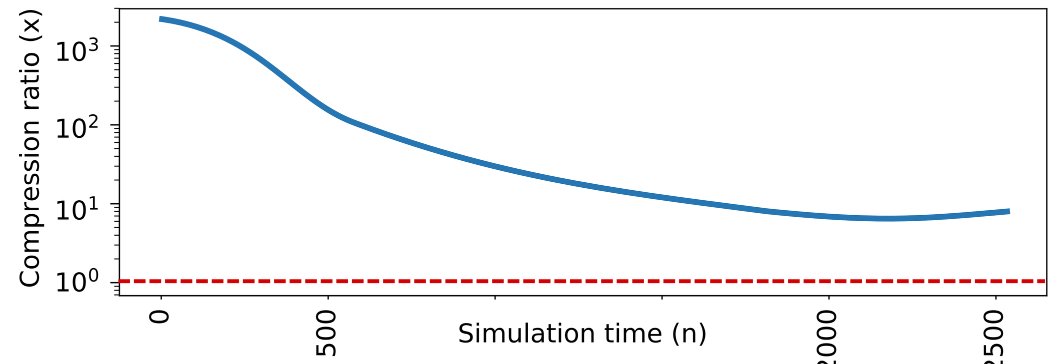


Figure 6: Compression ratios through simulation using ZFP at tolerance 10^{-3}

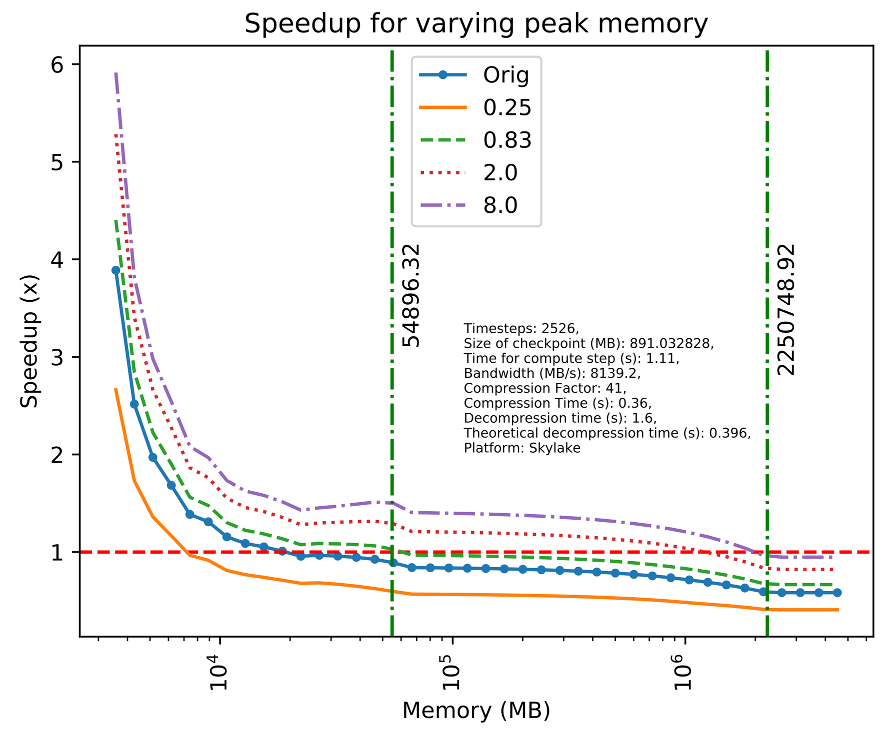


Figure 7: Speedup when using compression as compared to only recomputation for varying amount of memory

References

[1] R-E Plessix.
A review of the adjoint-state method for computing the gradient of a functional with geophysical applications.
Geophysical Journal International, 167(2):495–503, 2006.

[2] Jean Virieux and Stéphane Operto.
An overview of full-waveform inversion in exploration geophysics.
Geophysics, 74(6):WCC1–WCC26, 2009.

[3] Andreas Griewank and Andrea Walther.
Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation.
ACM Transactions on Mathematical Software (TOMS), 26(1):19–45, 2000.

[4] Navjot Kukreja, Jan Hückelheim, and Gerard J Gorman.
Backpropagation for long sequences: beyond memory constraints with constant overheads.
arXiv preprint arXiv:1806.01117, 2018.

[5] Navjot Kukreja, Jan Huckelheim, Mathias Louboutin, Kaiyuan Hou, Fabio Luporini, Paul Hovland, and Gerard Gorman.
Combining checkpointing and data compression for large scale seismic inversion.
arXiv preprint arXiv:1810.05268, 2018.

[6] Jan Hückelheim, Navjot Kukreja, Sri Hari Krishna Narayanan, Fabio Luporini, Gerard Gorman, and Paul Hovland.
Automatic differentiation for adjoint stencil loops.
arXiv preprint arXiv:1907.02818, 2019.

[7] Mathias Louboutin, Michael Lange, Fabio Luporini, Navjot Kukreja, Philipp A Witte, Felix J Herrmann, Paulius Velesko, and Gerard J Gorman.
Devito: an embedded domain-specific language for finite differences and geophysical exploration.
arXiv preprint arXiv:1808.01995, 2018.

Acknowledgements

Navjot Kukreja is funded by an Innovate UK grant.

Imperial College London