ParCFD'2021
32^{nd} International Conference on Parallel Computational Fluid Dynamics
May-17-19 2021, Nice FRANCE

ALGORITHMIC STRATEGIES FOR THE ACCELERATION OF FULL-WAVEFORM INVERSION

Andreas FICHTNER*, Solvi THRASTARSON* AND Dirk-Philip VAN HERWAARDEN*

* ETH Zürich
Department of Earth Sciences
Institute of Geophysics
Sonneggstrasse 5, 8092 Zürich, Switzerland
e-mail: andreas.fichtner@erdw.ethz.ch, web page: https://www.swp.ethz.ch

Key words: Geophysics, Seismology, Full-waveform inversion, Inverse Problems, Spectral-element method, High-performance computing

Abstract. We present a series of strategies that can reduce the computational cost of full-waveform inversion (FWI) by up to an order of magnitude for well-chosen but still widely relevant cases. These strategies include the use of wavefield-adapted meshed, stochastic gradient descent with dynamic mini-batches, and autotuning of stochastic FWI based on factorised quasi-Newton methods.

1 INTRODUCTION

The Earth is a 3D heterogeneous medium that causes elastic waves travelling through it to reflect, refract and scatter. The resulting wavefield complexity is a rich source of information that is still far from being completely exploited. A central role in our quest to improve images of the Earth's interior – or other inaccessible bodies – is being played by the comparatively recent development of full-waveform inversion (FWI) [1, 6]. Though the term is not uniquely defined, most FWI implementations share the numerical simulation of wave propagation through complex media, combined with adjoint techniques to compute derivatives.

In addition to its undeniable success in providing images with unprecedented detail, FWI is also notorious for its high computational cost. The No-Free-Lunch Theorem [4] precludes, from the outset, the existence of an FWI variant that is universally more efficient than other variants for all possible applications. This fundamental limitation motivates the development of strategies that accelerate FWI in well-defined but still reasonably broad classes of applications. Three of these strategies will be presented in this contribution: (1) The use of finite-element meshes that are adapted to *a priori* knowledge of wavefield geometry, (2) a stochastic gradient descent algorithm tuned towards seismological datasets in order to exploit their inherent redundancy, and (3) an auto-tuning approach based on quasi-Newton methods that may increase convergence of Hamiltonian Monte Carlo FWI by orders of magnitude.

2 WAVEFIELD-ADAPTED MESHES

The standard practice in numerical wavefield modelling is to use finite-element meshes adapted to the structure of the medium. This includes the alignment of element boundaries along discontinuities, and the use of larger elements in regions with higher wave speed.

The latter may be further reduced when prior knowledge about the geometry of the wavefield is available. Arguably the most prominent cases in this class of applications are radially symmetric Earth models that produce wavefields with exact azimuthal symmetry with respect to the source location. This symmetry enables the reduction of a 3D wave propagation problem to a quasi-2D problem, with corresponding savings in computational cost [5].

In practice, the requirement of exact azimuthal symmetry may be relaxed. Earth models that are smooth over length scales of several wavelengths, and which typically result from tomographic inversions, still produce wavefields with approximate azimuthal symmetry. This allows us to design finite-element meshes that are symmetric around the source location and have only a small number of elements in azimuthal direction [7, 8]. The latter depends on the smoothness of the medium and on the required solution accuracy, largely dictated by observed data errors.

Though wavefield-adapted meshes, when interpreted as physical entities, cannot resolve the adjoint field, the application of the discrete adjoint method [1] still provides the correct sensitivity kernels. The combination of wavefield-adapted forward and adjoint modelling enables FWI implementations that may require up to an order of magnitude lower computational resources [7].

3 STOCHASTIC GRADIENTS AND DYNAMIC MINI-BATCHES

Datasets used in regional- to global-scale FWI are typically characterised by significant redundancies, mostly resulting from the clustering of earthquakes in areas of elevated seismicity. Building on the stochastic gradient descent concept, widely used in machine learning, this redundancy can be exploited for the reduction of computational cost, without sacrificing the quality of the reconstructed model [9, 10].

Realising that waveform data from nearby earthquakes do not carry independent information, stochastic gradient descent starts with the selection of a suitable subset of N sources. Using adjoint techniques, the gradient of the subset, called a mini-batch, can be computed and used to obtain an updated model. Subsequently, a new mini-batch is selected quasi randomly, and the updating procedure is repeated. To ensure that a mini-batch is large enough to approximate the gradient of the complete dataset sufficiently well, its size can be adjusted. For this, we determine the number $n \leq N$ of sources within the mini-batch, needed to approach the mini-batch gradient to within a pre-defined error. The size of the mini-batch for the subsequent iteration is then set to a multiple of n.

Several synthetic and real-data inversions have shown that the dynamic mini-batch approach may reduce computational cost by up to 80 % for realistic 3D scenarios. This suggests that the method may be a game changer in FWI using earthquake datasets.

4 AUTOTUNING HAMILTONIAN MONTE CARLO

Especially in the context of local-scale seismic exploration, FWI is a highly nonlinear problem, characterised by the presence of local minima that result from the lack of low-frequency information combined with the unavailability of good initial models. Accounting for nonlinearity and multimodality in both model construction and uncertainty quantification typically requires the application of Monte Carlo sampling.

While often considered to be of reach, stochastic FWI is about to become feasible thanks to the development of Hamiltonian Monte Carlo (HMC) techniques that exploit derivative information [2, 3]. An outstanding advantage of HMC is its tunability, which enables very efficient variants of the algorithm. However, this advantage comes at the cost of actually having to tune in order to prevent inefficiency.

The most relevant tuning parameters in HMC are the mass matrix $\mathbf{M}$, the integration time step Δt, and the total integration time T. Though the optimal $\mathbf{M}$ is the local Hessian of the misfit surface, useful approximations can be found via quasi-Newton algorithms, and with the factorised L-BFGS algorithm in particular. Changes of $\mathbf{M}$ usually change the properties of the numerical integrator, thereby requiring adaptations of Δt and T. We perform these adaptations automatically by monitoring the HMC acceptance rate, and by making sure that it stays within a reasonable range, roughly between 0.6 and 0.95.

In a series of synthetic inversions, we are able to show that this autotuning approach has the potential to increase the effective sample size of an HMC chain by more than an order of magnitude, thereby further contributing to the practical applicability of HMC-based FWI.

5 CONCLUSIONS

We presented a collection of strategies for the acceleration of FWI in well-chosen but still widely relevant use cases. All of these are of purely algorithmic nature, and hence come in addition to soft- and hardware-based improvements. Synthetic and real-data inversions suggest reductions in computational cost by up to an order of magnitude. This implies that these strategies may enable FWI without the need of supercomputing resources and/or the use of significantly more data, potentially at higher frequency.

References

[1] A. Fichtner. *Full Seismic Waveform Modelling and Inversion.* Springer, Heidelberg, 2010.

[2] A. Fichtner, A. Zunino, and L. Gebraad. Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophys. J. Int.*, 216:doi:10.1093/gji/ggy496, 2019.

[3] L. Gebraad, C. Boehm, and A. Fichtner. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo. *J. Geophys. Res.*, 125, 2020.

[4] K. Mosegaard. *Limits to nonlinear inversion*, pages 11–21. Springer, Berlin, Heidelberg, 2012.

[5] T. Nissen-Meyer, M. van Driel, S. Stähler, K. Hosseini, S. Hempel, L. Auer, and A. Fournier. AxiSEM: broadband 3-D seismic wavefields in axisymmetric media. *Solid Earth*, 5:425–445, 2014.

[6] A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49:1259–1266, 1984.

[7] S. Thrastarson, M. van Driel, L. Krischer, C. Boehm, M. Afanasiev, D.-P. van Herwaarden, and A. Fichtner. Accelerating numerical wave propagation by wavefield adapted meshes, Part II: Full-waveform inversion. *Geophys. J. Int.*, 221:1591–1604, 2020.

[8] M. van Driel, C. Boehm, L. Krischer, and M. Afanasiev. Accelerating numerical wave-propagation using wavefield adapted meshes, Part I: Forward and adjoint modelling. *Geophys. J. Int.*, 211:1580–1590, 2020.

[9] D.-P. van Herwaarden, M. Afanasiev, S. Thrastarson, and A. Fichtner. Evolutionary full-waveform inversion. *Geophys. J. Int.*, 224:306–311, 2021.

[10] D.-P. van Herwaarden, C. Boehm, M. Afanasiev, S. Thrastarson, L. Krischer, M. van Driel, and A. Fichtner. Accelerated full-waveform inversion using dynamic mini-batches. *Geophys. J. Int.*, 221:1427–1438, 2020.