Ultra-low memory seismic inversion with randomized trace estimation

Mathias Louboutin[1], Felix J. Herrmann[1,2]
[1] School of Earth and Atmospheric Sciences, Georgia Institute of Technology
[2] School of Computational Science and Engineering, Georgia Institute of Technology

1   Summary:

Inspired by recent work on extended image volumes that lays the ground for randomized probing of extremely large seismic wavefield matrices, we present a memory frugal and computationally efficient inversion methodology that uses techniques from randomized linear algebra. By means of a carefully selected realistic synthetic example, we demonstrate that we are capable of achieving competitive inversion results at a fraction of the memory cost of conventional full-waveform inversion with limited computational overhead. By exchanging memory for negligible computational overhead, we open with the presented technology the door towards the use of low-memory accelerators such as GPUs.

2   Introduction

Wave-equation based seismic inversion has resulted in tremendous improvements in subsurface imaging over the past decade. However, the computational cost, and above-all memory cost, of these methods remains extremely high limiting their widespread adaptation. These memory costs stem from the requirement of the adjoint-state method [1, Tarantola [2]] to store (in memory, on disk, possibly compressed) the complete time history of the forward modeled wavefield prior to on-the-fly crosscorrelation with the time-reversed or adjoint wavefield. Unfortunately, the need to store these forward wavefields for each source may require allocation of up to terabytes of memory during high-frequency 3-D imaging. To tackle this memory requirement and to open the way to the use of low-memory accelerators such as GPUs, different methods have been proposed that balance memory usage with computational overhead to reduce the memory footprint. One of the earliest methods in this area is optimal checkpointing [3, Symes [4]], which has recently been extended to include wavefield compression [5]. Given available memory, optimal checkpointing reduces storage needs at the expense of having to recompute the forward wavefield. While the resulting memory savings can be significant, the computational overhead can be high (up to tens of extra wave-equations). Alternatively, by relying on time-reversibility of the (attenuation-free) wave-equation, McMechan [6], Mittet [7], Raknes and Weibull [8] proposed an approach where the forward wavefield is reconstructed during back propagation from values stored on the boundary. Even though these approaches have been applied successfully in practice, their implementation becomes elaborate especially in situations of complex wave physics such as in elastic or transverse-tilted anisotropic media.

Aside from the aforementioned methods that are in principle error free, except perhaps for methods that involve lossy compression, approximate methods have also been proposed to reduce the memory imprint of gradient calculations for wave-equation based imaging. The rationale behind these approximations is that by giving up accuracy one may gain computationally and/or memorywise. This was, for example, the main idea behind imaging with phase-encoded sources [9, Krebs et al. [10], Moghaddam et al. [11], Haber et al. [12], van Leeuwen and Herrmann [13]]. According to insights from stochastic optimization, there are strong arguments that accurate gradients are indeed unnecessary especially at the beginning of the optimization and as long as the approximate gradient equals the true gradient in expectation [14]. We make use of this fundamental observation and propose a new, simple to implement, alternative formulation that allows for approximate gradient calculations with a significantly reduced memory imprint. At first sight, our method may be somewhat reminiscent of memory-reducing methods based on the Fourier transform, which compute the Fourier transform

either on the fly [15, Witte et al. [16]] or in windows [17]. Our method is based on a randomized trace estimation technique instead. Like Fourier methods, which derive their advantage from computing the gradient for a relatively small number of Fourier modes, the proposed randomized algorithm collects compressed data during forward propagation. This leads to significant memory savings at a manageable computational overhead (similar if not cheaper than checkpointing [18]). However, compared to optimal checkpointing and boundary methods, Fourier techniques involve inaccurate gradients that may result in coherent artifacts during the inversion. Contrarily, our randomized linear algebra based method produces incoherent artifacts, easily handled by regularizations methods, with a computational overhead of at least half of that of Fourier methods thanks to real-valued arithmetic.

Inspired by recent work on randomized linear algebra for seismic inversion [19, Yang et al. [20]], we propose an alternative method based on random trace estimation. Unlike direct time subsampling [21], randomized trace estimation involves matrix probing with random vectors [22, Meyer et al. [23]] (vectors with random ±1's or Gaussian vectors) frequently employed by randomized algorithms, such as the randomized SVD [24]. These probings are computationally efficient because they only involve actions of the matrix on a small number of random vectors. This allows us to compress the time axes during forward propagation, greatly reducing memory usage in a computationally efficient way, yielding gradient calculations with a controllable error.

Our contributions are organized as follows. First, we introduce randomized trace estimation including a recently proposed orthogonalization step that improves its performance and accuracy. Next, we show that gradients calculated with the adjoint-state method for wave-equation based inversion can be approximated by random trace estimation. Compared to conventional gradient calculations, the approximated gradient leads to significant memory reductions and faster computation of certain imaging conditions. To further justify the proposed algorithm, we discuss how its computational and memory requirements compare to existing methods. We conclude by illustrating the advocacy of the proposed method on a 2D synthetic full-waveform inversion example.

3 Methodology

In this section, we briefly lay out the key components of our seismic inversion method that requires significantly less memory. We start by introducing stochastic estimates for the trace of a matrix, followed by how these estimates can be used to approximate the gradient of a time-domain formulation of the adjoint state method for the wave-equation. We conclude by providing estimates of memory use of the proposed method and how it compares to other memory reducing methods.

3.1 Randomized trace estimation

To address memory pressure of data-intensive applications, a new generation of randomized algorithms have been proposed. Contrary to classical deterministic techniques that aim for maximal accuracy, these methods are stochastic in nature and provide answers with a controllable error. Examples of these techniques are the randomized SVD [24, Yang et al. [20]] and randomized trace estimation [Avron, 23]. The latter was used to justify wave-equation inversions with random phase encoding [12, van Leeuwen and Herrmann [13]]. In this work, we also rely on randomized trace estimation but now to reduce the memory imprint of wave-equation based inversion. At its heart, randomized trace estimation [22, Meyer et al. [23]] derives from the following approximation of the identity $\mathbf{I}$:

$$
\begin{aligned}
\operatorname{tr}(\mathbf{A}) = \operatorname{tr}\left(\mathbf{A}\mathbb{E}\left[\mathbf{z}\mathbf{z}^\top\right]\right) &= \mathbb{E}\left[\operatorname{tr}\left(\mathbf{A}\mathbf{z}\mathbf{z}^\top\right)\right] \\
&= \mathbb{E}\left[\mathbf{z}^\top \mathbf{A}\mathbf{z}\right] \approx \frac{1}{r}\sum_{i=1}^{r}\left[\mathbf{z}_i^\top \mathbf{A}\mathbf{z}_i\right] = \frac{1}{r}\operatorname{tr}\left(\mathbf{Z}^\top \mathbf{A}\mathbf{Z}\right)
\end{aligned}
\tag{1}
$$

where the $\mathbf{z}_i$'s are the random probing vectors for which $\mathbb{E}(\mathbf{z}^\top\mathbf{z}) = 1$ and $\mathbb{E}$ is the stochastic expectation operator. The above estimator is unbiased (exact in expectation) and converges to the true trace of the matrix $\mathbf{A}$, i.e. $\operatorname{tr}(A) = \sum_i \mathbf{A}_{ii}$, with an error that decays with r and without access to the entries of $\mathbf{A}$. Only actions of $\mathbf{A}$ on the probing vectors are needed and we will exploit this property and the factored form of the matrix $\mathbf{A}$ in gradient calculations for wave-equation based inversion. Motivated by recent work [23, Graff-Kray et al. [25]] we will also employ a partial qr factorization [26] that approximates the range of the matrix $\mathbf{A}$—i.e., we approximate the trace with probing vectors $[\mathbf{Q}, \sim] = \operatorname{qr}(\mathbf{A}\mathbf{Z})$ where $\mathbf{Z}$ is a Rademacher random matrix of ±1.

3.2 Approximate gradient calculations

While one may argue that inaccurate gradients need to be avoided at all times, approximations to the gradient are actually quite common. For instance, wave-equation based inversion with phase-encoded sources derives from a similar approximation but this time on the data misfit objective. Haber et al. [12] showed that this type of approximate inversion is an instance of random trace estimation. In addition, possible artifacts can easily be removed by adding regularization, e.g. by imposing the TV-norm constraint on the model [27, Peters et al. [28]] or by including curvelet-domain sparsity constraints on Gauss-Newton updates [29]. Let us consider the standard adjoint-state FWI problem, which aims to minimize the misfit between recorded field data and numerically modeled synthetic data [1, Tarantola [2], Virieux and Operto [30], Louboutin et al. [31], Louboutin et al. [32]]. In its simplest form, the data misfit objective for this problem reads

$$\underset{\mathbf{m}}{\text{minimize}} \; \frac{1}{2}||\mathbf{F}(\mathbf{m};\mathbf{q}) - \mathbf{d}_{\text{obs}}||_2^2 \tag{2}$$

where $\mathbf{m}$ is a vector with the unknown physical model parameter (squared slowness in the isotropic acoustic case), $\mathbf{q}$ the sources, $\mathbf{d}_{\text{obs}}$ the observed data and $\mathbf{F}$ the forward modeling operator. This data misfit is typically minimized with gradient-based optimization methods such as gradient descent [33] or Gauss-Newton [29]. While the presented approach carries over to arbitrary complex wave physics, we derive our memory reduced gradient approximation for the isotropic acoustic case where the gradient for a single source $\delta\mathbf{m}$ can be written as

$$\delta\mathbf{m} = \sum_t \ddot{\mathbf{u}}[t]\mathbf{v}[t] \tag{3}$$

where $\mathbf{u}[t], \mathbf{v}[t]$ are the vectorized (along space) full-space forward and adjoint solutions of the forward and adjoint wave-equation at time index t. The symbol ¨ represents second-order time derivative. To arrive at a form where randomized trace estimation can be used, we write the above zero-lag crosscorrelation over time as the trace of the outer product for each space index $\mathbf{x}$ separately. By using the dot product property, $\sum \mathbf{x}_i \mathbf{y}_i = \mathbf{x}^\top \mathbf{y} = \text{tr}(\mathbf{x}\mathbf{y}^\top)$, in combination with Equation 1, we approximate the gradient via random-trace estimation—i.e.,

$$\delta\mathbf{m}[\mathbf{x}] = \text{tr}\left(\ddot{\mathbf{u}}[t,\mathbf{x}]\mathbf{v}[t,\mathbf{x}]^\top\right) \approx \frac{1}{r}\,\text{tr}\left((\mathbf{Q}^\top\ddot{\mathbf{u}}[\mathbf{x}])(\mathbf{v}[\mathbf{x}]^\top\mathbf{Q})\right) \tag{4}$$

where parenthesis were added to show that the matrix-vector products between the wavefields and the probing matrix $\mathbf{Q}$ can be computed independently. We use this property to arrive at our proposed ultra-low memory unbiased approximation of the gradient summarized in Algorithm 1 below. This algorithm runs for each space index with the same probing matrix $\mathbf{Q}$. Instead of storing the wavefield during the forward pass, e.g. via checkpointing, the sum of the probed (with $\mathbf{Q}^\top$) wavefield is for each space index accumulated (line 2) in the variable $\bar{\mathbf{u}}[\mathbf{x}]$. Because of the probing, we only need to store $N \times r$ with $r \ll n_t$ samples in $\bar{\mathbf{u}}$ instead of $N \times n_t$ with n_t the number of timesteps. As we will show, r can be very small compared to n_t, leading to significant memory savings. Similarly, during the back propagation pass (lines 4-7), we accumulate for each spatial index $\bar{\mathbf{v}}$. After finishing the back propagation iterations, the gradient is calculated by computing the trace of the outer product of the accumulated probed wavefields (line 8). To avoid forming an unnecessarily large matrix for the outer product in line 8, we simply compute the sum over the probing size of the pointwise product of $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ in practice. Before deriving a practical scheme for FWI based on this probing scheme, let us first discuss its memory use and how it compares to other known approaches to reduce the memory imprint of wave-equation based inversion.

3.3 Memory estimates

From Equation 4, we can easily estimate the memory imprint of our method compared to conventional FWI. For completeness, we also consider other mainstream low memory methods: optimal checkpointing [3, Symes [4], Kukreja et al. [5]], boundary methods [6, Mittet [7], Raknes and Weibull [8]], and DFT methods [17, Sirgue et al. [15], Witte et al. [16]]. This memory overview generalizes to other wave-equations and imaging conditions easily as our method generalizes to any time-domain adjoint-state method. We estimate the memory requirements for a three-dimensional domain with $N = N_x \times N_y \times N_z$ grid points and n_t time steps. Conventional FWI requires storing the full time-space forward wavefield to compute the gradient. This requirement leads to a memory

Algorithm 1 Approximate gradient calculation with random trace estimation
___
0. **for t=2:nt-1**   # forward propagation
1. $\mathbf{u}[t+1] = f(\mathbf{u}[t], \mathbf{u}[t-1], \mathbf{m}, \mathbf{q}[t])$
2. $\bar{\mathbf{u}}[\mathbf{x}, \cdot] \mathrel{+}= \mathbf{Q}[\cdot, t]^{\top} \ddot{\mathbf{u}}[t, \mathbf{x}]$
3. **end for**
4. **for t=nt:-1:1**   # back propagation
5. $\mathbf{v}[t-1] = f^{\top}(\mathbf{v}[t], \mathbf{v}[t+1], \mathbf{m}, \delta \mathbf{d}[t])$
6. $\bar{\mathbf{v}}[\cdot, \mathbf{x}] \mathrel{+}= \mathbf{v}[t, \mathbf{x}] \mathbf{Q}[\cdot, t]$
7. **end for**
8. output: $\frac{1}{r} \operatorname{tr}(\bar{\mathbf{u}}\, \bar{\mathbf{v}}^{\top})$
___

requirement of $N \times n_t$ floating point values. Our method, on the other hand, for r probing vectors (i.e $\mathbf{Q} \in \mathbb{R}^{n_t \times r}$), requires only $N \times r$ floating point values during each of the forward and backward passes for a total of $2 \times N \times r$ values. The memory reduction factor is, therefore, $\frac{n_t}{2r}$. This memory reduction is similar to computing the gradient with $\frac{r}{2}$ Fourier modes. We summarize the memory usage compared to other state-of-the-art algorithms in table 1.

	FWI	DFT	Probing	Optimal checkpointing	Boundary reconstruction
Compute	0	$\mathcal{O}(2r) \times n_t \times N$	$\mathcal{O}(r) \times n_t \times N$	$\mathcal{O}(log(n_t)) \times N \times n_t$	$n_t \times N$
Memory	$N \times n_t$	$2r \times N$	$r \times N$	$\mathcal{O}(10) \times N$	$n_t \times N^{\frac{2}{3}}$

Table 1: Memory estimates and computational overhead of different seismic inversion methods for n_t time steps and N grid points.

It is worth noting that unlike the other methods in the table, boundary reconstruction methods tend to have stability issues for more complex physics, in particular with physical attenuation making it ill-suited for real-world applications. As stated in the introduction, our method closely follows the computational and memory cost of Fourier methods by a factor of two related to real versus complex arithmetic. We also show that unlike checkpointing or boundary methods, neither the memory nor computational overhead depends on the number of time steps, therefore our method offers improved scalability.

3.4   Imaging conditions

Aside from these clear advantages regarding memory use, the proposed approximation scheme also has computational advantages when imposing more elaborate imaging conditions such as the inverse scattering imaging condition [34, Witte et al. [35]] for RTM or wavefield separation [36] for FWI. In most cases, these imaging condition can be expressed as linear operators that only act on the spatial dimensions of the wavefields and not along time. Because these operators are linear, we can factor these operators out and directly apply them to the probed wavefields consisting of r reduced time steps rather than to every time steps with $n_t \gg r$. We can achieve this by making use of the following identity (the same applies to $\mathbf{v}$):

$$\mathbf{Q}^{\top}\left(\mathbf{D}_x \mathbf{u}[\cdot, \mathbf{x}]\right) = \mathbf{D}_x \left(\mathbf{Q}^{\top} \mathbf{u}[\cdot, \mathbf{x}]\right), \tag{5}$$

which holds as long as the linear imaging condition only acts along the space directions and not along time. Because $r \ll n_t$ this can lead to significant computational savings especially in the common situation where imposing imaging conditions becomes almost as expensive as solving the wave-equation itself.

3.5   Choice of the probing matrix Q

While strictly random, e.g. random ± 1 as in Rademacher or Gaussians [22] an extra orthogonalization step (via a $\mathtt{qr}$ factorization on random probings $\mathbf{A}\mathbf{Z}$) allows us to capture the range of $\mathbf{A}$ [23, Graff-Kray et al. [25]], which leads faster decay of the error as a function of r. This error in the gradient is due to "cross talk"—i.e. $\mathbf{Z}\mathbf{Z}^{\top} \neq \mathbf{I}$. Unfortunately, we do not have easy access to $\mathbf{A}$ during the approximate gradient calculations outlined in Algorithm 1. Moreover, orthogonalizing each spatial

gridpoint separately would be computationally infeasible. Despite these complications, we argue that we can still get a reasonable approximation of the range by random probing the observed data organized as a matrix for each source experiment—i.e., we have

$$[\mathbf{Q}, \sim] = \mathrm{qr}(\mathbf{AZ}) \quad \text{with} \quad \mathbf{A} = \mathbf{D}_{\mathrm{obs}}\mathbf{D}_{\mathrm{obs}}^{\top} \tag{6}$$

where the observed data vector $\mathbf{d}_{\mathrm{obs}}$ for each shot is shaped into a matrix along the time and lumped together receiver coordinates. Since observed data contains information on the temporal characteristics of the wavefields, we argue that this outer product can serve as a proxy for the time characteristics of the wavefield everywhere. In Figure 1, we demonstrate the benefits of the additional orthogonalization step by comparing the outer products of the Rademacher probing matrix $\mathbf{Z}$, the restricted Fourier matrix $\mathbf{F}$, and orthogonalized probing vectors $\mathbf{Q}$ as a function of increasing r. The following observations can be made. First, as expected the "cross-talk", i.e. amplitudes away from the diagonal, becomes smaller when r increases, which is to be expected for all cases. However, we observe also that the outer product converges faster to the identity for both the Rademacher and qr factored case while coherent artifact remain with Fourier due to the truncation. Second, because of the orthogonalization the artifacts for the outer product of $\mathbf{Q}$ are much smaller and this should improve the accuracy of the gradient at the expense of a relatively minor cost of carrying our a qr factorization for each shot record. Finally, compared to the Fourier basis, in combination with source spectrum informed frequency sampling, our probing factors are informed by estimates of the range of the sample covariance kernel spanned by the traces in each shot record.
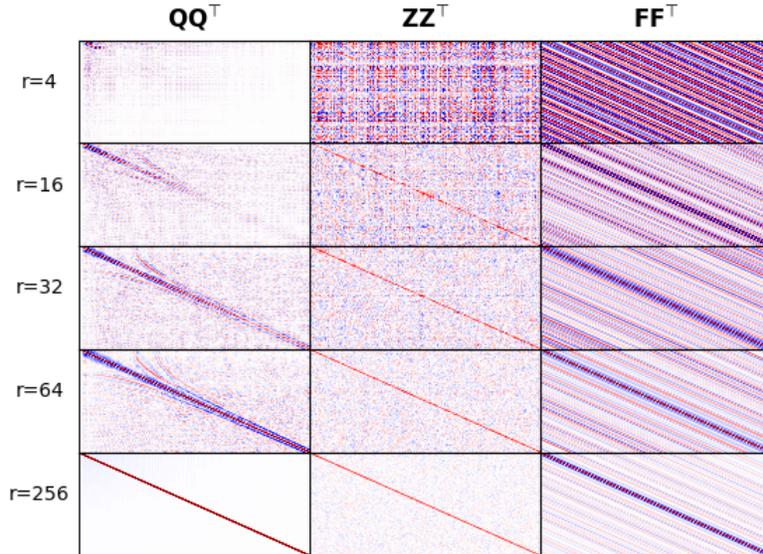
Figure 1: Probing vector for varying probing size.

4   Example

We illustrate our method on the 2D overthrust model and compare our inversion results to conventional FWI and on-the-fly DFT [15, Witte et al. [16]]. We consider a 20km by 5km 2D slice of the well-known overthrust model. The dataset consists of 97 sources 200m apart at 50m depth. Each shot contains between 127 and 241 ocean bottom nodes 50m apart at 500m depth for a maximum offset of 6km. The data is modeled with an 8Hz Ricker wavelet and 3sec recording. We show the true model and the initial background model with the inversion results in Figure 2. We ran 20 iterations of Spectral Projected Gradient (SPG, gradient descent with box constraints [37]) with 20 randomly selected shots per iteration [38] in all cases. We can clearly see from Figure 2 that our probed gradient allows the inversion to carry towards a good velocity estimate. As theoretically expected, for few probing vectors, we do not converge since our approximation is not accurate enough. However, we start to obtain a result comparable to the true model with as few as 16 probing vectors. Additionally, this result could easily be improved by adding constraints as regularization [27, Peters et al. [28]]. On the other hand, we can also see in Figure 2 that for an equivalent memory cost, on-the-fly DFT fails to converge to an acceptable result for any number of frequencies, most likely due to the coherent

artifacts that stem from the DFT. These results could also be improved with constraints or with a better choice of selected frequencies. We finally compare the three vertical traces highlighted in black to detail the accuracy of the inverted velocity plotted in Figure 2. These traces show that our probed inversion result is in the vicinity of results obtained with standard FWI, which itself is close to the true model.
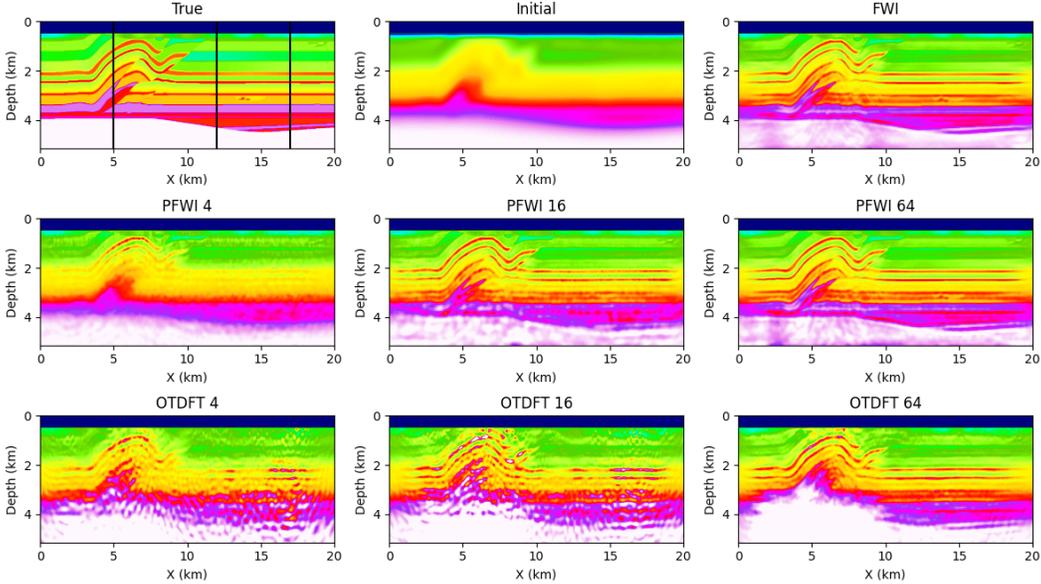
Figure 2: Probed versus On-the-fly DFT FWI on the 2D overthrust model with equivalent memory imprints.
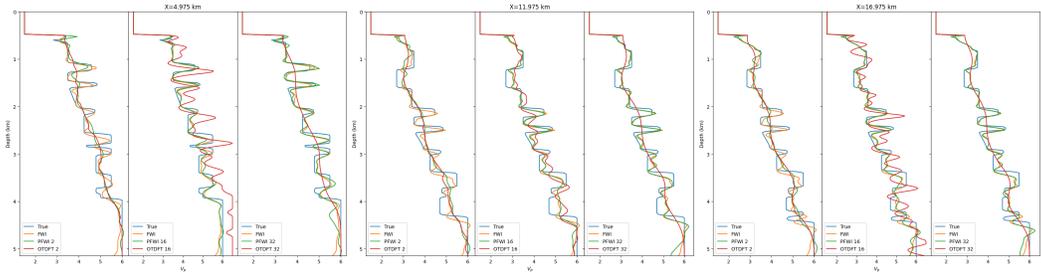
Figure 3: Vertical trace comparison between the OTDFT and probed FWI.

Our implementation and examples are open sourced at TimeProbeSeismic.jl and extends our Julia inversion framework JUDI.jl. Our code is also designed to generalize to 3D and more complicated physics as supported by Devito [39, Luporini et al. [40]].

5 Discussion and Conclusions

We introduced a randomized trace estimation technique to drastically reduce the memory footprint of wave-equation based inversion. We achieved this result at a computational overhead similar smaller than that of Fourier-based methods. However, compared to these methods our approach is simpler and produces less coherent crosstalk. Aside from the elegance and simplicity of the well-established technique of randomized trace estimation our proposed approach derives its performance from probing vectors that approximate the range of the data's sample covariance operator. We successfully demonstrated our randomized scheme on a realistic 2D full-waveform example where our method outperforms Fourier based methods. In future work, we plan to extend our methodology to high-frequency inversion and to per offset or even per trace probing vectors instead of per shot. We will also test this method on more complex wave physics.

6

6 Acknowledgement

7 References

References

[1] J. L. Lions. *Optimal control of systems governed by partial differential equations*. Springer-Verlag Berlin Heidelberg, 1st edition, 1971. ISBN 978-3-642-65026-0.

[2] Albert Tarantola. Inversion of seismic reflection data in the acoustic approximation. *GEO-PHYSICS*, 49(8):1259, 1984. doi: 10.1190/1.1441754. URL +http://dx.doi.org/10.1190/1.1441754.

[3] Andreas Griewank and Andrea Walther. Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Softw.*, 26(1):19–45, March 2000. ISSN 0098-3500. doi: 10.1145/347837.347846. URL http://doi.acm.org/10.1145/347837.347846.

[4] Symes. Reverse time migration with optimal checkpointing. *GEOPHYSICS*, 72(5):SM213–SM221, 2007. doi: 10.1190/1.2742686. URL http://library.seg.org/doi/abs/10.1190/1.2742686.

[5] N. Kukreja, J. Hückelheim, M. Louboutin, J. Washbourne, P. H. J. Kelly, and G. J. Gorman. Lossy checkpoint compression in full waveform inversion. *Geoscientific Model Development Discussions*, 2020:1–26, 2020. doi: 10.5194/gmd-2020-325. URL https://gmd.copernicus.org/preprints/gmd-2020-325/.

[6] G. A. McMechan. Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting*, 31(3):413–420, 1983. ISSN 1365-2478. doi: 10.1111/j.1365-2478.1983.tb01060.x. URL http://dx.doi.org/10.1111/j.1365-2478.1983.tb01060.x.

[7] Rune Mittet. Implementation of the kirchhoff integral for elastic waves in staggered-grid modeling schemes. *GEOPHYSICS*, 59(12):1894–1901, 1994. doi: 10.1190/1.1443576. URL http://dx.doi.org/10.1190/1.1443576.

[8] Espen Birger Raknes and Wiktor Weibull. Efficient 3d elastic full-waveform inversion using wavefield reconstruction methods. *Geophysics*, 81(2):R45–R55, 2016. ISSN 0016-8033. doi: 10.1190/geo2015-0185.1. URL http://geophysics.geoscienceworld.org/content/81/2/R45.

[9] Louis A. Romero, Dennis C. Ghiglia, Curtis C. Ober, and Scott A. Morton. Phase encoding of shot records in prestack migration. *GEOPHYSICS*, 65(2):426–436, 2000. doi: 10.1190/1.1444737. URL https://doi.org/10.1190/1.1444737.

[10] Jerome R. Krebs, John E. Anderson, David Hinkley, Ramesh Neelamani, Sunwoong Lee, Anatoly Baumstein, and Martin-Daniel Lacasse. Fast full-wavefield seismic inversion using encoded sources. *GEOPHYSICS*, 74(6):WCC177–WCC188, 2009. doi: 10.1190/1.3230502. URL https://doi.org/10.1190/1.3230502.

[11] Peyman P. Moghaddam, Henk Keers, Felix J. Herrmann, and Wim A. Mulder. A new optimization approach for source-encoding full-waveform inversion. *GEOPHYSICS*, 78(3):R125–R132, 2013. doi: 10.1190/geo2012-0090.1. URL https://doi.org/10.1190/geo2012-0090.1.

[12] Eldad Haber, Kees van den Doel, and Lior Horesh. Optimal design of simultaneous source encoding. *Inverse Problems in Science and Engineering*, 23(5):780–797, 2015. doi: 10.1080/17415977.2014.934821. URL https://doi.org/10.1080/17415977.2014.934821.

[13] Tristan van Leeuwen and Felix J. Herrmann. Fast waveform inversion without source-encoding. *Geophysical Prospecting*, 61(s1):10–19, 2013. doi: https://doi.org/10.1111/j.1365-2478.2012.01096.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2478.2012.01096.x.

[14] Michael P. Friedlander and Mark Schmidt. Hybrid deterministic-stochastic methods for data fitting. *SIAM Journal on Scientific Computing*, 34(3):A1380–A1405, 2012. doi: 10.1137/110830629. URL https://doi.org/10.1137/110830629.

[15] L. Sirgue, J.T. Etgen, U. Albertin, and S. Brandsberg-Dahl. System and method for 3d frequency domain waveform inversion based on 3d time-domain forward modeling, May 2010. URL http://www.google.ca/patents/US7725266. US Patent 7,725,266.

[16] Philipp A. Witte, Mathias Louboutin, Fabio Luporini, Gerard J. Gorman, and Felix J. Herrmann. Compressive least-squares migration with on-the-fly fourier transforms. *Geophysics*, 84(5): R655–R672, 08 2019. doi: 10.1190/geo2018-0490.1. URL https://slim.gatech.edu/Publications/Public/Journals/Geophysics/2019/witte2018cls/witte2018cls.pdf. (Geophysics).

[17] Kurt T. Nihei and Xiaoye Li. Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (td–psd). *Geophysical Journal International*, 169(3):1069–1078, 2007. doi: 10.1111/j.1365-246X.2006.03262.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.2006.03262.x.

[18] Philipp A. Witte, Mathias Louboutin, Henryk Modzelewski, Charles Jones, James Selvage, and Felix J. Herrmann. An event-driven approach to serverless seismic imaging in the cloud. *IEEE Transactions on Parallel and Distributed Systems*, 31(9):2032–2049, 03 2020. doi: 10.1109/TPDS.2020.2982626. URL https://slim.gatech.edu/Publications/Public/Journals/IEEETPDS/2020/witte2019TPDedas/witte2019TPDedas.html. (IEEE Transactions on Parallel and Distributed Systems).

[19] Tristan van Leeuwen, Rajiv Kumar, and Felix J. Herrmann. Enabling affordable omnidirectional subsurface extended image volumes via probing. *Geophysical Prospecting*, 65(2):385–406, 03 2017. doi: 10.1111/1365-2478.12418. URL https://slim.gatech.edu/Publications/Public/Journals/GeophysicalProspecting/2016/vanleeuwen2015GPWEMVA/vanleeuwen2015GPWEMVA.pdf. (Geophysical Prospecting).

[20] Mengmeng Yang, Marie Graff, Rajiv Kumar, and Felix J. Herrmann. Low-rank representation of omnidirectional subsurface extended image volumes. *Geophysics*, 86(3):1–41, 1 2021. doi: 10.1190/geo2020-0152.1. URL https://slim.gatech.edu/Publications/Public/Journals/Geophysics/2021/yang2020lrpo/Paper_final.html. (Geophysics).

[21] Mathias Louboutin and Felix J. Herrmann. Time compressively sampled full-waveform inversion with stochastic optimization. In *SEG Technical Program Expanded Abstracts*, pages 5153–5157, 10 2015. doi: 10.1190/segam2015-5924937.1. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/SEG/2015/louboutin2015SEGtcs/louboutin2015SEGtcs.html. (SEG, New Orleans).

[22] Haim Avron and Sivan Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *J. ACM*, 58(2), April 2011. ISSN 0004-5411. doi: 10.1145/1944345.1944349. URL https://doi.org/10.1145/1944345.1944349.

[23] Raphael A. Meyer, Cameron Musco, Christopher Musco, and David P. Woodruff. Hutch++: Optimal Stochastic Trace Estimation. *arXiv e-prints*, art. arXiv:2010.09649, October 2020.

[24] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011. doi: 10.1137/090771806. URL https://doi.org/10.1137/090771806.

[25] Marie Graff-Kray, Rajiv Kumar, and Felix J. Herrmann. Low-rank representation of omnidirectional subsurface extended image volumes, 2017. URL https://slim.gatech.edu/Publications/Public/Conferences/SINBAD/2017/Fall/graff2017SINBADFlrp/graff2017SINBADFlrp.pdf.

[26] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.

[27] Ernie Esser, Lluís Guasch, Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann. Total-variation regularization strategies in full-waveform inversion. *SIAM Journal on Imaging Sciences*, 11(1):376–406, 2018. doi: 10.1137/17M111328X. URL https://slim.gatech.edu/Publications/Public/Journals/SIAMJournalOnImagingSciences/2018/esser2016tvr/esser2016tvr.pdf. (SIAM Journal on Imaging Sciences).

[28] Bas Peters, Brendan R. Smithyman, and Felix J. Herrmann. Projection methods and applications for seismic nonlinear inverse problems with multiple constraints. *Geophysics*, 84(2):R251–R269, 02 2018. doi: 10.1190/geo2018-0192.1. URL https://slim.gatech.edu/Publications/Public/Journals/Geophysics/2018/peters2018pmf/peters2018pmf.html. (Geophysics).

[29] Xiang Li, Ernie Esser, and Felix J. Herrmann. Modified gauss-newton full-waveform inversion explained–-why sparsity-promoting updates do matter. *Geophysics*, 81(3):R125–R138, 05 2016. doi: 10.1190/geo2015-0266.1. URL https://slim.gatech.edu/Publications/Public/Journals/Geophysics/2016/li2015GEOPmgn/li2015GEOPmgn.pdf. (Geophysics).

[30] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *GEOPHYSICS*, 74(5):WCC1–WCC26, 2009. doi: 10.1190/1.3238367. URL http://library.seg.org/doi/abs/10.1190/1.3238367.

[31] Mathias Louboutin, Philipp A. Witte, Michael Lange, Navjot Kukreja, Fabio Luporini, Gerard Gorman, and Felix J. Herrmann. Full-waveform inversion - part 1: forward modeling. *The Leading Edge*, 36(12):1033–1036, 12 2017. doi: 10.1190/tle36121033.1. URL https://slim.gatech.edu/Publications/Public/Journals/TheLeadingEdge/2017/louboutin2017fwi/louboutin2017fwi.html. (The Leading Edge).

[32] Mathias Louboutin, Philipp A. Witte, Michael Lange, Navjot Kukreja, Fabio Luporini, Gerard Gorman, and Felix J. Herrmann. Full-waveform inversion - part 2: adjoint modeling. *The Leading Edge*, 37(1):69–72, 1 2018. doi: 10.1190/tle37010069.1. URL https://slim.gatech.edu/Publications/Public/Journals/TheLeadingEdge/2018/louboutin2017fwip2/louboutin2017fwip2.html. (The Leading Edge).

[33] 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. ISSN 1365-246X. doi: 10.1111/j.1365-246X.2006.02978.x. URL http://dx.doi.org/10.1111/j.1365-246X.2006.02978.x.

[34] N. D. Whitmore and Sean Crawley. *Applications of RTM inverse scattering imaging conditions*, pages 1–6. 2012. doi: 10.1190/segam2012-0779.1. URL https://library.seg.org/doi/abs/10.1190/segam2012-0779.1.

[35] P.A. Witte, M. Yang, and F.J. Herrmann. Sparsity-promoting least-squares migration with the linearized inverse scattering imaging condition. 2017(1):1–5, 2017. ISSN 2214-4609. doi: https://doi.org/10.3997/2214-4609.201701125. URL https://www.earthdoc.org/content/papers/10.3997/2214-4609.201701125.

[36] Faqi Liu, Guanquan Zhang, Scott A. Morton, and Jacques P. Leveille. An effective imaging condition for reverse-time migration using wavefield decomposition. *GEOPHYSICS*, 76(1):S29–S39, 2011. doi: 10.1190/1.3533914. URL https://doi.org/10.1190/1.3533914.

[37] Mark Schmidt, Ewout Berg, Michael Friedlander, and Kevin Murphy. Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. 5:456–463, 16–18 Apr 2009. URL http://proceedings.mlr.press/v5/schmidt09a.html.

[38] Aleksandr Y. Aravkin, Michael P. Friedlander, Felix J. Herrmann, and Tristan van Leeuwen. Robust inversion, dimensionality reduction, and randomized sampling. *Mathematical Programming*, 134(1):101–125, 08 2012. doi: 10.1007/s10107-012-0571-6. URL http://www.springerlink.com/content/35rwr101h5736340/.

[39] M. Louboutin, M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman. Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. *Geoscientific Model Development*, 12(3):1165–1187, 2019. doi: 10.5194/gmd-12-1165-2019. URL https://www.geosci-model-dev.net/12/1165/2019/.

[40] Fabio Luporini, Mathias Louboutin, Michael Lange, Navjot Kukreja, Philipp Witte, Jan Hückelheim, Charles Yount, Paul H. J. Kelly, Felix J. Herrmann, and Gerard J. Gorman. Architecture and performance of devito, a system for automated stencil computation. *ACM Trans. Math. Softw.*, 46(1), April 2020. ISSN 0098-3500. doi: 10.1145/3374916. URL https://doi.org/10.1145/3374916.