Parsimonious time-domain truncated-Newton method in FWI thanks to Fourier-domain and "full scattered field" approximation

Peng Yong, Univ. Grenoble Alpes, ISTerre; Romain Brossier, Univ. Grenoble Alpes, ISTerre; Ludovic Métivier, Univ. Grenoble Alpes, CNRS, LJK*

SUMMARY

In order to exploit Hessian information in Full Waveform Inversion (FWI), matrix-free truncated-Newton method can be used. In such a method, the Hessian-vector product can be estimated thanks to four wavefield simulations. Different from frequency-domain implementation, it is not feasible to store the whole time-dependent wavefields in time-domain FWI. The widely used recomputation strategy leads to computationally intensive tasks. In this work, we propose an affordable parsimonious alternative to compute such Hessian-vector product, thanks to a Fourier-domain compression and a "full scattered field" approximation of the wavefields involved. Numerical tests on multi-parameter Hessian-vector product construction illustrates that the proposed method can significantly reduce computational time within acceptable accuracy.

INTRODUCTION

With the success of application of mono-parameter FWI to field data in relatively simple geological environments (Warner et al., 2013; Operto et al., 2015), application of FWI to multi-parameter and complex environments attracts more and more attention to academia and industry (Operto and Miniussi, 2018; Plessix, 2019). In such a frame, the Hessian matrix plays an important role when involved in the non-linear optimization, in order to accelerate convergence rate, consider second order scattering and mitigate cross-talk between different parameters (Pratt et al., 1998; Fichtner and Trampert, 2011; Métivier et al., 2013; Pan et al., 2016). The truncated-Newton strategy relies on applying the Hessian information in a matrix-free fashion during the descent direction computation through a linear conjugate-gradient-based system and has been shown to be a powerful tool in FWI (Métivier et al., 2013, 2015; Pan et al., 2018; Yang et al., 2018).

The key ingredients of the truncated-Newton method are the construction of the gradient vector and Hessian-vector product via first and second order adjoint-state methods (Métivier et al., 2013). In frequency-domain, as storing single frequency-domain wavefields in memory is relatively cheap even for large scale problem, the cost of Hessian-vector product is not so expensive, in particular when direct solvers are involved (Métivier et al., 2013). However, it becomes much more expensive in the time-domain since storing a whole time-dependent wavefields in memory is challenging, and recomputation strategies have to be implemented leading to high computational cost (Yang et al., 2018). Source encoding (Castellanos et al., 2015) and subsampling shot strategy (Matharu and Sacchi, 2019) can be applied to reduce the computational cost in "coarse-grained" way. In this work, we propose a parsimonious approach for time-domain FWI in a "fine-grained" way, using all the shots for gradient, but relying on some approximations to significantly decrease the computation effort of Hessian-vector product. Considering that frequency-domain wavefields are relatively cheap to store, and relying on the same kind of "on-the-fly" Fourier transform than the ones proposed for gradient building with phase-sensitive detection (Nihei and Li, 2007) or discrete Fourier transform (Sirgue et al., 2008), we can extract the Fourier-domain components of the wavefields on-the-fly and save them for several frequencies due to the band-limited nature of seismic data. Therefore, we can approximate the Hessian-vector product thanks to a Fourier-domain version with few frequencies. By doing that, we successfully avoid the computationally intensive task of reconstructing the first and second-order incident wavefields, in particular when viscous media are involved (Yang et al., 2016b).

In addition to this Fourier-domain approximation, a second improvement relies on the fact that the second-order incident and adjoint wavefields can be interpreted as first-order Born scattering wavefields generated from interactions between model perturbations and first-order incident and adjoint wavefields. Such equivalent first-order Born modeling is challenging to implement in 3D, as source terms of the second-order incident and adjoint wavefields are volumetric, time-dependent and related to the first-order wavefields. The physical meaning of the second-order incident and adjoint wavefields is related to linear wavefield change with a medium perturbation (Schuster, 2017). Relying on this physical interpretation and for small perturbations, we could approximate such second-order wavefields with the difference between one frequency-domain wavefield computed in unperturbed model and one computed in perturbed model.

In this work, we first introduce the brief theory of truncated-Newton algorithm for FWI with adjoint-state method. Then, we develop our proposed approximation formulation with on-the-fly Fourier-domain compression and the "full scattered field" approximation frame. An application to the synthetic 2D Valhall case is then presented.

METHODOLOGY

Truncated-Newton method in time-domain

FWI updates the model by minimizing the difference between the observed and predicted data, which can be formulated as a PDE-constrained optimization problem of the form

$$\min_{\mathbf{m}} \chi(\mathbf{m}) = \frac{1}{2}\int_0^T ||R\mathbf{w}-\mathbf{d})||^2 \mathrm{d}t, \qquad (1)$$
$$s.t. \quad A(\mathbf{m})\mathbf{w}-\mathbf{s}=0, \quad \mathbf{w}(\mathbf{x},t)|_{t=0}=0,$$

where $\mathbf{m} \in \mathbf{R}^n$ is the model parameters of interest in model space, $\mathbf{d} := \mathbf{d}(\mathbf{x}_r,t)$ is the observed data at receiver location $\mathbf{x}_r$, R is receiver sampling operator. $A(\mathbf{m})$ represents the forward modeling operator and $\mathbf{s}$ is the source term. In the framework of truncated-Newton method, the optimization algorithm can

10.1190/segam2020-3426296.1

be described by the iteration

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \alpha^k \Delta\mathbf{m}^k, \tag{2}$$

where α^k represents step length at the k-th iteration. $\Delta\mathbf{m}^k$ is the model update direction and can be obtained by solving (or approximating) the Newton equation

$$H(\mathbf{m}^k)\Delta\mathbf{m}^k = -\nabla\chi(\mathbf{m}^k). \tag{3}$$

$\nabla\chi(\mathbf{m})$ and $H(\mathbf{m}) = \nabla^2\chi(\mathbf{m})$ are the gradient and Hessian matrix. With the adjoint-state method, the gradient of time-domain FWI can be given by

$$\nabla\chi(\mathbf{m}) = \langle\lambda, \frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\mathbf{w}\rangle_T = \int_0^T \mathrm{d}t\lambda^\dagger \frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\mathbf{w}, \tag{4}$$

where λ is the adjoint wavefield obtained by solving the first-order adjoint-state equation

$$A^\dagger(\mathbf{m})\lambda = R^\dagger(\mathbf{d} - R\mathbf{w}), \quad \lambda(\mathbf{x},t)|_{t=T} = 0. \tag{5}$$

It is important to point out that, different from the initial condition in the equation (1), the first-order adjoint-state equation contains a final condition, which indicates back-propagating the data residuals (Plessix, 2006). To construct Hessian-vector product $H(\mathbf{m})\mathbf{r}$ with second-order adjoint-state method, two extra equations are required to be solved:

$$A(\mathbf{m})\mathbf{u} = -\sum_{m_i\in\mathbf{m}} r_i \frac{\partial A(\mathbf{m})}{\partial m_i}\mathbf{w}, \quad \mathbf{u}(\mathbf{x},t)|_{t=0} = 0. \tag{6}$$

$$A^\dagger(\mathbf{m})\mu = -R^\dagger R\mathbf{u} - \sum_{m_i\in\mathbf{m}} r_i \left(\frac{\partial A(\mathbf{m})}{\partial m_i}\right)^\dagger \lambda, \quad \mu(\mathbf{x},t)|_{t=T} = 0, \tag{7}$$

where $\mathbf{r} \in \mathbf{R}^n$ is the given vector. $\mathbf{u}$ and μ are the second-order incident and adjoint wavefields, for which the source term depends on first-order fields. The full Newton Hessian-vector product can then be expressed by

$$\begin{aligned} H(\mathbf{m})\mathbf{r} = &\langle\mathbf{u}, \frac{\partial A^\dagger(\mathbf{m})}{\partial\mathbf{m}}\lambda\rangle_T + \langle\mu, \frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\mathbf{w}\rangle_T \\ &+ \langle\lambda, \left(\frac{\partial^2 A(\mathbf{m})}{\partial\mathbf{m}^2}\mathbf{w}\right)\mathbf{r}\rangle_T. \end{aligned} \tag{8}$$

We could make $\partial^2 A(\mathbf{m})/\partial\mathbf{m}^2 = 0$ by selecting suitable parameterization when anisotropic parameters are not reconstructed. After constructing the Hessian-vector product with the chosen parameterization scheme, we can easily obtain Hessian-vector product of other parameterization schemes using the chain rule (Yang et al., 2018). Gauss-Newton Hessian, a semi-positive approximation of full Newton Hessian is widely used in practice, which is given by

$$H(\mathbf{m})\mathbf{r} = \langle\mu, \frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\mathbf{w}\rangle_T, \tag{9}$$

where μ is defined by

$$A^\dagger(\mathbf{m})\mu = -R^\dagger R\mathbf{u}, \quad \mu(\mathbf{x},t)|_{t=T} = 0. \tag{10}$$

In order to build the gradient vector (equation (4)) and Hessian-vector product (equations (8) and (9)), the forward and the backward wavefields have to be accessed simultaneously for the cross-correlation computation. As it is usually not feasible to store all the time-dependent incident wavefields for realistic application, check-pointing and recomputation strategies are often implemented (Griewank and Walther, 2000; Symes, 2007; Yang et al., 2016a). The main computationally expansive steps for time-domain full Newton Hessian-vector product construction are shown in Algorithm 1.

Algorithm 1: full Newton Hessian-vector product construction in the time domain

for *it* = 1 *to nt* **do**
  update incident fields $\mathbf{w}(\mathbf{x},t)$ and $\mathbf{u}(\mathbf{x},t)$;
end
for *it* = *nt to* 1 **do**
  update adjoint fields $\lambda(\mathbf{x},t)$ and $\mu(\mathbf{x},t)$;
  reconstruct incident fields $\mathbf{w}(\mathbf{x},t)$ and $\mathbf{u}(\mathbf{x},t)$;
  build the Hessian-vector product $H(\mathbf{m})\mathbf{r}$ in time-domain;
end

On-the-fly Fourier transform

In order to avoid the computationally intensive tasks of recomputing the incident wavefields for Hessian-vector product in the time-domain, we consider using discrete Fourier transform to obtain frequency-domain wavefields, and we propose to approximate the Hessian-vector product by a frequency-domain version with only a small number of frequencies. The on-the-fly DFT of the forward wavefield $\mathbf{w}(\mathbf{x},t)$ is given by

$$\widetilde{\mathbf{w}}(\mathbf{x},f) = \sum_{k=1}^{nt} exp(-2\pi i f k\Delta t)\mathbf{w}(\mathbf{x},k\Delta t)\Delta t, \tag{11}$$

where $i = \sqrt{-1}$, $\widetilde{\mathbf{w}}(\mathbf{x},f)$ denotes frequency-domain wavefield. Δt is the sampling interval and is determined by the Nyquist theorem.

After obtaining the four frequency-domain wavefields, we can construct Hessian-vector product in the frequency-domain with two terms (the anisotropic parameters are not considered in this work).

$$H(\mathbf{m})\mathbf{r} = 2\int_0^{f_{max}} \mathscr{R}\left(\widetilde{\mathbf{u}}^\dagger \frac{\partial\widetilde{A}^\dagger(\mathbf{m})}{\partial\mathbf{m}}\widetilde{\lambda} + \widetilde{\mu}^\dagger \frac{\partial\widetilde{A}(\mathbf{m})}{\partial\mathbf{m}}\widetilde{\mathbf{w}}\right)\mathrm{d}f, \tag{12}$$

where $\widetilde{A}(\mathbf{m})$ is the forward modeling operator in frequency-domain. $\mathscr{R}$ means taking the real part of complex number.

Full scattered field approximation

Let us consider the forward equation for computing the field $\mathbf{w}_1$ governed by

$$A(\mathbf{m}+\mathbf{r})\mathbf{w}_1 = \mathbf{s}, \tag{13}$$

where $\mathbf{r}$ is a small model perturbation. The wavefield $\mathbf{w}_1$ can be written as

$$\mathbf{w}_1 = \mathbf{w} + \delta\mathbf{w}, \tag{14}$$

where the wavefield $\mathbf{w}$ is the solution of

$$A(\mathbf{m})\mathbf{w} = \mathbf{s}, \tag{15}$$

and $\delta\mathbf{w}$ can be seen as the "full scattered field" generated from the model perturbation $\mathbf{r}$ added to $\mathbf{m}$. This "full scattered field" can be decomposed as

$$\delta\mathbf{w} = \delta\mathbf{w}_1 + \delta\mathbf{w}_2 + ..., \tag{16}$$

10.1190/segam2020-3426296.1

where $\delta\mathbf{w}_j$ ($j = 1, 2, ...$) is the jth-order Born scattered wavefield, which is recursively given by

$$A(\mathbf{m})\delta\mathbf{w}_j = -\sum_{m_i \in \mathbf{m}} r_i \frac{\partial A(\mathbf{m})}{\partial m_i}(\delta\mathbf{w}_{j-1}). \tag{17}$$

For $j = 1$, wavefield in the source term $\delta\mathbf{w}_0$ is defined as $\mathbf{w}$ (Schuster, 2017). One can recognize that, as $j = 1$, equation (17) is similar to equation (6), meaning that $\mathbf{u}$ is the first-order Born wavefield $\delta\mathbf{w}_1$ caused by the model perturbation $\mathbf{r}$.

From the previous development, assuming that $\mathbf{r}$ is small enough to neglect second and higher order scattered terms, we can approximate $\delta\mathbf{w}_1 \approx \delta\mathbf{w}$, and therefore compute $\mathbf{u}$ as the difference between $\mathbf{w}_1$ and $\mathbf{w}$, which gives

$$\begin{cases} A(\mathbf{m})\mathbf{w}(\mathbf{x},t) = \mathbf{s}, \\ A(\mathbf{m}+\mathbf{r})\mathbf{w}_1(\mathbf{x},t) = \mathbf{s}, \\ \mathbf{u}(\mathbf{x}) \approx \mathbf{w}_1(\mathbf{x}) - \mathbf{w}(\mathbf{x}). \end{cases} \tag{18}$$

By implementing this "full scattered field" approximation of $\mathbf{u}(\mathbf{x})$ in the Fourier domain, we only need to solve one wave equation for each Hessian-vector product, considering that $\widetilde{\mathbf{w}}(\mathbf{x}, f)$ is already known and stored.

For the second-order adjoint wavefields μ in the full Newton approximation, the same strategy can be used, but leads to a small inaccuracy because of the $R^\dagger R\mathbf{u}$ in the perturbed model instead of the originally unperturbed model:

$$\begin{cases} A^\dagger(\mathbf{m})\lambda(\mathbf{x},t) = R^\dagger(\mathbf{d} - R\mathbf{w}), \\ A^\dagger(\mathbf{m}+\mathbf{r})\lambda_1(\mathbf{x},t) = R^\dagger(\mathbf{d} - R\mathbf{w}) - R^\dagger R\mathbf{u}, \\ \mu(\mathbf{x}) \approx \lambda_1(\mathbf{x}) - \lambda(\mathbf{x}). \end{cases} \tag{19}$$

By using the Fourier domain approximation of the wavefields, we end-up with the Algorithm 2 for the main steps of our "full scattered field" approximation.

Algorithm 2: full Newton Hessian-vector product with frequency-domain "full scattered field" approximation

$\widetilde{\mathbf{w}}(\mathbf{x}, f)$ and $\widetilde{\lambda}(\mathbf{x}, f)$ are computed in the gradient construction
for $it = 1$ *to* nt **do**
  update incident fields by solving $A(\mathbf{m}+\mathbf{r})\mathbf{w}_1(\mathbf{x},t) = \mathbf{s}$;
  update the frequency-domain incident field $\widetilde{\mathbf{w}}_1(\mathbf{x}, f)$
end
Approximate second-order incident wavefield
 $\widetilde{\mathbf{u}}(\mathbf{x}, f) \approx \widetilde{\mathbf{w}}_1(\mathbf{x}, f) - \widetilde{\mathbf{w}}(\mathbf{x}, f)$;
Approximate adjoint source for second-order adjoint
 wavefield $R^\dagger R\mathbf{u} \approx R^\dagger R\mathbf{w}_1 - R^\dagger R\mathbf{w}$
for $it = nt$ *to* 1 **do**
  update adjoint fields by solving
   $A^\dagger(\mathbf{m}+\mathbf{r})\lambda_1(\mathbf{x},t) = R^\dagger(\mathbf{d} - R\mathbf{w}) - R^\dagger R\mathbf{u}$;
  update the frequency-domain adjoint field $\widetilde{\lambda}_1(\mathbf{x}, f)$;
end
Approximate second-order incident wavefield
 $\widetilde{\mu}(\mathbf{x}, f) \approx \widetilde{\lambda}_1(\mathbf{x}, f) - \widetilde{\lambda}(\mathbf{x}, f)$;
build the Hessian-vector product $H(\mathbf{m})\mathbf{r}$ with $\widetilde{\mathbf{w}}$, $\widetilde{\lambda}$, $\widetilde{\mathbf{u}}$ and $\widetilde{\mu}$

Computational complexity analysis

The main computationally intensive steps of Gauss and full Newton Hessian-vector product construction are the wavefield simulations and the time-domain cross-correlation steps.

By analyzing algorithms 1, we can observe that the original time-domain formulation leads to 6 wavefield simulations per Hessian-vector product: 2 forward fields, 2 backward fields and 2 recomputation of forward fields backward in time or with checkpointing strategies.

Analyzing algorithm 2 shows that our frequency-domain "full scattered field" approximation only needs 2 wavefield simulations per Hessian-vector product: 1 forward and 1 backward fields, and no any recomputation steps which can be quite intensive in particular in viscous media.

NUMERICAL EXPERIMENTS

2D multi-parameter Valhall model shown in Figure 1 is defined on a regular grid ($nz = 281$, $nx = 704$). The spatial interval is 12.5 m. A fixed-spread acquisition with 32 equally spaced sources on the surface and 489 equally spaced receivers placed at the top, left and right sides with interval of 25 m. The synthetic data is generated with second order in time and fourth order in space finite-difference modeling. Source function is Ricker wavelet with peak frequency of 5 Hz. The time discretization step is set to 1.5 ms. The maximum frequency considered here is about 12.5 Hz. 20 frequencies are extracted in the frequency-domain equally sampled from 0-12.5 Hz for Hessian-vector product.

The forward modeling operator $A(\mathbf{m})$ discretizes an acoustic VTI wave equation (Yang et al., 2018). The computational times for constructing one Hessian-vector product via the two approaches, namely the time-domain implementation of Yang et al. (2018) and the frequency-domain "full scattered field" approximation of algorithm 2 are listed in Table 1. As expected, the proposed approximation allows to reduce significantly the recomputation tasks and therefore reduce the "time-to-solution" to obtain Hessian-vector product. It has to be noted that in attenuative media, the recomputation effort would be even bigger for the time-domain implementation, our approximation involve only forward-time propagation which are much less demanding.

Figures 2 and 3 show the Gauss-Newton and full Newton Hessian-vector product provided by the two schemes for velocity, density and attenuation parameters. Here, the negative gradient is used as the vector $\mathbf{r}$ for Hessian-vector product. As Hessian plays the role of convolution operator, $H\mathbf{r}$ is a blurred version of the negative gradient. It can be noted that the proposed method provides solutions very close to the reference version in the time-domain with 20 frequencies considered here.

Table 1: Comparison of elapsed time per shot for Gauss-Newton and full Newton Hessian-vector product construction via two methods.

	Gauss-Newton	full Newton
time-domain	306 s	570 s
proposed approximation	84 s	91 s

10.1190/segam2020-3426296.1

CONCLUSION

A parsimonious approach for the truncated-Newton method with a time-domain solver has been proposed via a "full scattered field" approximation in Fourier domain. This makes possible to reduce significantly the computational effort of Hessian-vector product, while maintaining an acceptable accuracy. Thanks to Fourier-domain compression, we can avoid the computationally intensive tasks of recomputing the incident wavefields. Full scattered field can effectively approximate the first-order Born scattered field when perturbation is small, which allows us updating second-order incident and adjoint wavefields without solving first-order incident and adjoint equations. Future work will tackle the number of frequencies required in the approximation and extension to 3D case.

ACKNOWLEDGMENTS

This study was partially funded by the SEISCOPE consortium (https://seiscope2.osug.fr), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC, SISPROBE and TOTAL. This study was granted access to the HPC resources of CIMENT infrastructure (https://ciment.ujf-grenoble.fr) and CINES/IDRIS/TGCC under the allocation 046091 made by GENCI.

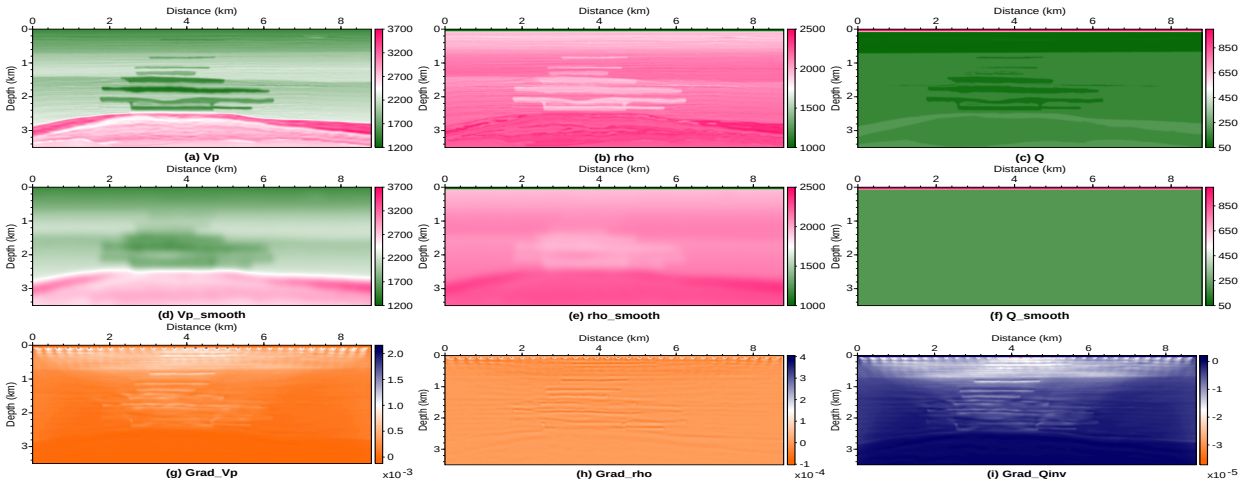

Figure 1: *P*-wave velocity (left column), density (middle column) and attenuation (right column) models. True models (first row), initial models (second row) and gradients in initial model (third row).

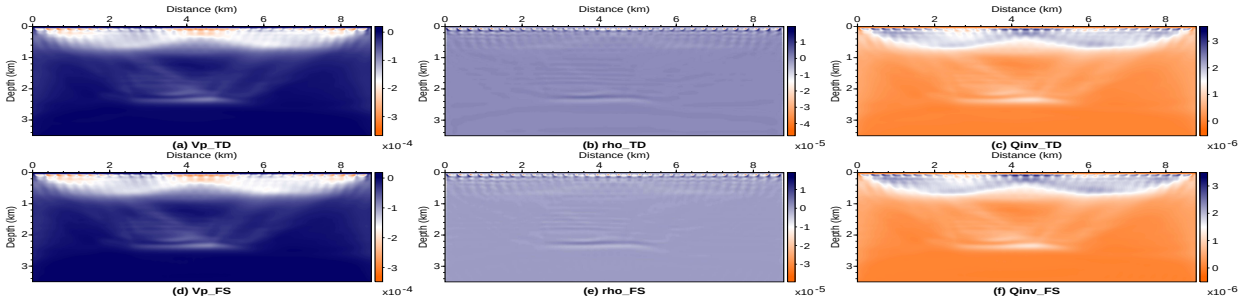

Figure 2: Gauss-Newton Hessian-vector product: *P*-wave velocity (left column), density (middle column) and attenuation (right column). Time-domain implementation (first row), "full scattered field" approximation in the frequency-domain (second row).

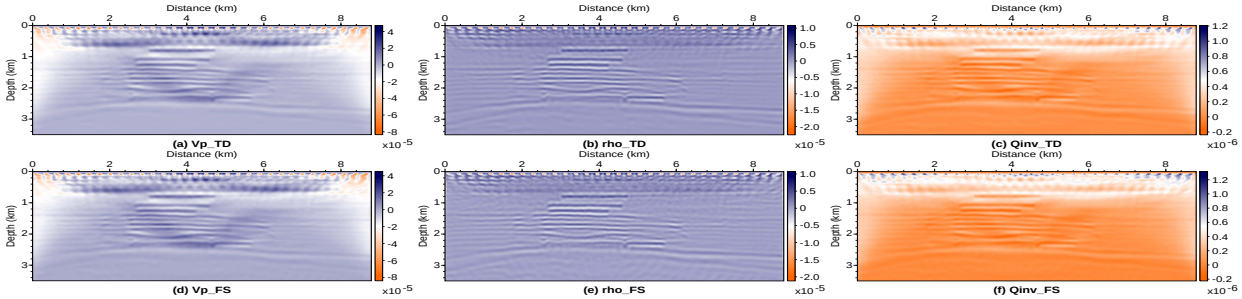

Figure 3: Full Newton Hessian-vector product: *P*-wave velocity (left column), density (middle column) and attenuation (right column). Time-domain implementation (first row), "full scattered field" approximation in the frequency-domain (second row).

10.1190/segam2020-3426296.1

REFERENCES

Castellanos, C., L. Métivier, S. Operto, R. Brossier, and J. Virieux, 2015, Fast full waveform inversion with source encoding and second-order optimization methods: Geophysical Journal International, **200**, 720–744, doi: https://doi.org/10.1093/gji/ggu427 .

Fichtner, A., and J. Trampert, 2011, Hessian kernels of seismic data functionals based upon adjoint techniques: Geophysical Journal International, **185**, 775–798, doi: https://doi.org/10.1111/j.1365-246X.2011.04966.x .

Griewank, A. and A. Walther, 2000, Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation: ACM Transactions on Mathematical Software, **26**, 19–45, doi: https://doi.org/10.1145/347837.347846.

Matharu, G., and M. Sacchi, 2019, A subsampled truncated-newton method for multi-parameter full waveform inversion: Geophysics, **84**, no. 3, R333–R340, doi: https://doi.org/10.1190/geo2018-0624.1.

Métivier, L., R. Brossier, S. Operto, and J. Virieux, 2015, Acoustic multi-parameter FWI for the reconstruction of P- wave velocity, density and attenuation: preconditioned truncated Newton approach: 85th Annual Meeting-New Orleans, Expanded Abstracts, 1198–1203, doi: https://doi.org/10.1190/segam2015-5875643.1.

Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full Waveform Inversion and the truncated Newton method: SIAM Journal on Scientific Computing, **35**, B401–B437, doi: https://doi.org/10.1137/120877854.

Nihei, K. T., and X. Li, 2007, Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD): Geophysical Journal International, **169**, 1069–1078, doi: https://doi.org/10.1111/j.1365-246X.2006.03262.x.

Operto, S., and A. Miniussi, 2018, On the role of density and attenuation in 3D multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea: Geophysical Journal International, **213**, 2037–2059, doi: https://doi.org/10.1093/gji/ggy103.

Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3- D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation: Geophysical Journal International, **202**, 1362–1391, doi: https://doi.org/10.1093/gji/ggv226.

Pan, W., K. A. Innanen, G. F. Margrave, M. C. Fehler, X. Fang, and J. Li, 2016, Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion: Geophysics, **81**, no. 5, R275–R291, doi: https://doi.org/10.1190/geo2015-0594.1.

Pan, Y., S. Schaneng, T. Steinweg, and T. Bohlen, 2018, Estimating s-wave velocities from 3d 9-component shallow seismic data using local rayleigh-wave dispersion curves – a field study: Journal of Applied Geophysics, **159**, 532–539, doi: https://doi.org/10.1016/j.jappgeo.2018.09.037.

Plessix, R., 2019, Should we move towards multi-parameter elastic inversions?: 81st EAGE Conference and Exhibition, 1–3.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503, doi: https://doi.org/10.1111/j.1365-246X.2006.02978.x.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362, doi: https://doi.org/10.1046/j.1365-246X.1998.00498.x.

Schuster, G. T., 2017, Seismic inversion: SEG.

Sirgue, L., J. T. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: Proceedings 70th EAGE, Conference and Exhibition, Roma, Italy, Expanded Abstracts, F022.

Symes, W. W., 2007, Reverse time migration with optimal check pointing: Geophysics, **72**, no. 5, SM213–SM221, doi: https://doi.org/10.1190/1.2742686.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, no. 2, R59–R80, doi: https://doi.org/10.1190/geo2012-0338.1.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016a, Checkpointing-assisted reverse forward simulation: an optimal recomputation method for fwi and rtm: SEG Technical Program, Expanded Abstracts, 1089–1093, doi: https://doi.org/10.1111/j.1365-246X.2010.04848.x.

Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018, A Time-domain preconditioned truncated Newton approach to multiparameter visco-acoustic full waveform inversion: SIAM Journal on Scientific Computing, **40**, B1101–B1130.

Yang, P., R. Brossier, and J. Virieux, 2016b, Wavefield reconstruction from significantly decimated boundaries: Geophysics, **80**, no. 5, T197–T209, doi: https://doi.org/10.1190/geo2015-0711.1.