ADAPT-THEN-COMBINE FULL WAVEFORM INVERSION FOR DISTRIBUTED SUBSURFACE IMAGING IN SEISMIC NETWORKS

Ban-Sok Shin and Dmitriy Shutin

Institute of Communications and Navigation, German Aerospace Center
E-Mail: {ban-sok.shin, dmitriy.shutin}@dlr.de

ABSTRACT

We consider the problem of distributed subsurface imaging in seismic receiver networks. This problem is particularly relevant for future planetary exploration missions where multi-agent networks shall autonomously reconstruct a subsurface based on network-wide measurements. To this end, we propose a distributed implementation of the full waveform inversion (FWI) for distributed imaging of subsurfaces in seismic networks. In particular, we show that the gradient of FWI is equivalent to the sum of locally computed gradients. To obtain estimates of the global gradient and subsurface model at each receiver we employ the adapt-then-combine technique that relies on data exchange among neighboring receivers only. Numerical evaluations show that the proposed distributed FWI performs close to its centralized version for different source-receiver constellations.

Index Terms— Distributed imaging, full waveform inversion, seismic imaging, inverse problems

1. INTRODUCTION

Full waveform inversion (FWI) is one of the state-of-the-art techniques for subsurface imaging in seismic surveys. Initially, it has been proposed in the time-domain using the acoustic wave equation by [1] and has been applied to geophysical exploration to obtain high-resolution images of subsurfaces [2, 3, 4, 5]. Recently, it has also gained much interest from the medical imaging community for ultrasound imaging of the human brain or breast [6, 7]. In FWI, a least-squares cost between observed and synthesized data is minimized with respect to a model parameter of the subsurface, e.g., the seismic wave velocity. The objective of FWI is then to iteratively improve the model of the subsurface by reducing the residual between the observed and synthesized receiver data. The strength of FWI lies in the use of the whole measured waveforms of the receivers and the exploitation of a wave equation that properly describes the underlying physics of wave propagation. However, FWI needs to solve a non-convex optimization problem with respect to the model parameter, which requires careful selection of a numerical optimization method.

FWI has been nearly exclusively considered with a centralized inversion of the measurement data where the complete data of the receivers are available at a central entity. However, we envision the use of a multi-agent network without such a central entity that autonomously explores a certain subsurface e.g. in future planetary exploration missions [8]. For such an application, a distributed imaging technique is required. Here, each agent shall obtain a global subsurface model by interacting with other agents in the network such that dependency on a central entity is avoided.

Distributed inference methods have been thoroughly studied during the last decade, see e.g. [9, 10] and references therein. Nev-

ertheless, they have not yet been largely adopted by geophysical imaging techniques such as FWI. In [11] a decentralized FWI is proposed using the alternating direction method of multipliers while in [12] the authors describe a distributed reverse time migration. However, both works assume a central master node that fuses model updates of all receivers and therefore, do not enable for a fully distributed imaging of the subsurface. In [13, 14] distributed methods are proposed that exploit the physics of the wave equation for their inference task. However, both works focus on the localization of sources only.

In this paper, we propose a method that estimates the global FWI subsurface model in a distributed fashion at each receiver by exchanging data among neighboring receivers only. Then each receiver itself is able to compute a local model of the subsurface. Furthermore, we derive that the global gradient in FWI is equivalent to the sum of locally computed gradients. We show that the global gradient and the model can be estimated at each receiver using the adapt-then-combine method. We verify our proposed method by numerical simulations and show that the models obtained at each receiver are indeed close to those of a centralized FWI.

2. SEISMIC NETWORK MODEL

We consider a seismic network of J geophone receivers placed uniformly on a line for a two-dimensional exploration task. The topology of the seismic network is described via a graph $\mathcal{G} = \{\mathcal{J}, \mathcal{E}\}$ with a set of nodes $\mathcal{J} = \{1, 2, \ldots, J\}$ and a set of edges $\mathcal{E} = \{(j, i) | j, i \in \mathcal{J}, j \neq i\}$. We assume that the graph $\mathcal{G}$ is undirected and strongly connected, i.e., each receiver can be reached by any other receiver in the network over multiple hops. Furthermore, we define a neighborhood set $\mathcal{N}_j$ for each receiver j that contains those receivers that are directly connected to receiver j and the receiver j itself. Each neighbor contained in $\mathcal{N}_j$ can exchange data with receiver j and vice versa. For the exploration task, we assume N_s different shot locations on the surface and a total measurement time of T per shot $s = \{1, 2, \ldots, N_s\}$. Each receiver j has a fixed Cartesian position denoted by $\boldsymbol{x}_j = (x_j, z_j)$ with the x- and z-coordinate.

3. BRIEF REVIEW OF FULL WAVEFORM INVERSION

The objective of FWI is to minimize a cost function with respect to a specific model parameter of the wave equation to obtain a model that is close to the true one. The used cost function usually consists of some norm of a residual between measured seismic data and estimated, synthetic measurements generated from the estimated model. By minimizing this residual with respect to the parameters of the propagating wave, FWI is then able to reconstruct the original subsurface model. Here, we consider the P-wave velocity as the pa-

rameter of interest in the subsurface and we aim at recovering the velocity distribution over the considered area. A typical choice for the cost function in FWI is a least-squares cost between original seismic measurements $\boldsymbol{d}_{\text{obs}}$ and synthetically generated measurements $\boldsymbol{d}_{\text{syn}}$ as a function of the model parameter m, cf. [2]:

$$\mathcal{L}(m) = \frac{1}{2}||\boldsymbol{d}_{\text{obs}} - \boldsymbol{d}_{\text{syn}}(m)||_2^2. \tag{1}$$

In our case, the parameter $m = m(\boldsymbol{x})$ is a function that describes the wave velocity at a certain position $\boldsymbol{x} = (x, z)$ of the explored subsurface $\mathcal{X} \in \mathbb{R}^2$. The variable $\boldsymbol{d}_{\text{obs}}$ is a stacked vector of the measured, discrete time signals of each receiver while $\boldsymbol{d}_{\text{syn}}$ contains estimated, synthetic measurements based on the model m. Assuming a total number of N_T time samples per measurement and receiver j, $\boldsymbol{d}_{\text{obs}}$ and $\boldsymbol{d}_{\text{syn}}$ are vectors of dimension $JN_T \times 1$. The objective is to minimize the misfit between observed and generated measurements with respect to the model parameter m. To model the propagation of seismic waves in the subsurface we use the acoustic wave equation:

$$\frac{1}{m(\boldsymbol{x})^2}\frac{\partial^2 u(\boldsymbol{x}, t)}{\partial t^2} - \frac{\partial^2 u(\boldsymbol{x}, t)}{\partial \boldsymbol{x}^2} = f(\boldsymbol{x}, t), \tag{2}$$

where $f(\boldsymbol{x}, t)$ is the seismic source term and $u(\boldsymbol{x}, t)$ is the seismic wavefield at position $\boldsymbol{x}$ and time t, respectively, that describes the wave propagation through the subsurface. Since (2) describes the propagation of P-waves only, conversion into other wave types such as S-waves at subsurface interfaces is not considered here. To reduce unwanted reflections at the borders of the considered computational domain we include absorbing boundary layers in the numerical simulation. To minimize (1) iteratively with respect to the velocity m we need to compute the gradient of $\mathcal{L}(m)$. A common approach is to use the *adjoint-state method* to obtain the gradient, cf. [15]. Accordingly, the gradient can then be computed via

$$w(\boldsymbol{x}) = \frac{\partial \mathcal{L}(m)}{\partial m} = -\frac{2}{m^3(\boldsymbol{x})}\sum_{s=1}^{N_s}\int_0^T q_s(\boldsymbol{x}, T-t)\frac{\partial^2 u_s(\boldsymbol{x}, t)}{\partial t^2}dt. \tag{3}$$

The field $u_s(\boldsymbol{x}, t)$ is the forward-solved wavefield for the shot s using the model m. The field $q_s(\boldsymbol{x}, T-t)$ is called *adjoint wavefield* is generated by injecting the residual data $\boldsymbol{d}_{\text{obs}} - \boldsymbol{d}_{\text{syn}}(m)$ at the receiver locations as a source term and solving the wave equation backwards in time for a specific shot s. It can be shown that $q_s(\boldsymbol{x}, T-t)$ satisfies the following wave equation [15]:

$$\left(\frac{1}{m^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \Delta\right)q_s(\boldsymbol{x}, t) = \sum_{j\in\mathcal{J}}(\boldsymbol{d}_{\text{obs},j} - \boldsymbol{d}_{\text{syn},j}(m))\delta(\boldsymbol{x} - \boldsymbol{x}_j), \tag{4}$$

where Δ is the spatial Laplace operator and $\delta(\cdot)$ is the Delta-function. The time correlation of both fields under the integral in (3) is well known as the *imaging condition* or *migration* [16]. With the gradient $w(\boldsymbol{x})$ at hand, a gradient-descent method can be employed to iteratively minimize the cost $\mathcal{L}(m)$. To employ numerical optimization, we discretize the continuous space $\mathcal{X}$ into grid cells assuming using e.g. a finite-difference method. If we then collect velocity values of the model over all grid cells into a vector $\boldsymbol{m}^{[k]}$ and likewise all gradient values into a vector $\boldsymbol{w}^{[k]}$ at iteration k, we can update the current model via

$$\boldsymbol{m}^{[k+1]} = \boldsymbol{m}^{[k]} + \alpha^{[k]}\boldsymbol{w}^{[k]}. \tag{5}$$

The step size $\alpha^{[k]} > 0$ is usually found using a line search method to guarantee a reduction of the cost $\mathcal{L}(m)$ in each iteration k. To improve the convergence behavior of FWI, other methods such as the conjugate gradient method or L-BFGS can be used.

4. DISTRIBUTED FULL WAVEFORM INVERSION

The FWI described in Section 3 is a centralized processing algorithm since the adjoint wavefield $q_s(\boldsymbol{x}, t)$ per shot s is solved using the complete residual data of all J receivers. Based on $q_s(\boldsymbol{x}, t)$ the global gradient $\boldsymbol{w}$ is computed. To enable a distributed FWI the global gradient needs to be estimated locally at each receiver such that a local update of the velocity model per receiver is possible. In the following, we employ the adapt-then-combine technique to obtain an estimate of the global gradient $\boldsymbol{w}$ and the global model $\boldsymbol{m}$ at each receiver j. First, we find that the cost $\mathcal{L}(m)$ in (1) can be separated over the receivers:

$$\mathcal{L}(m) = \frac{1}{2}||\boldsymbol{d}_{\text{obs}} - \boldsymbol{d}_{\text{syn}}(m)||_2^2 = \sum_{j\in\mathcal{J}}\mathcal{L}_j(m) \tag{6}$$

with $\mathcal{L}_j(m) = \frac{1}{2}||\boldsymbol{d}_{\text{obs},j} - \boldsymbol{d}_{\text{syn},j}(m)||_2^2$ being the local cost function and $\boldsymbol{d}_{\text{obs},j}$ and $\boldsymbol{d}_{\text{syn},j}(m)$ being the measured and generated data vector at receiver j of dimension $N_T \times 1$, respectively. Then we compute the gradient of each local cost $\mathcal{L}_j(m)$ with respect to the model parameter m. Analogous to (3) we obtain the following gradient per receiver j:

$$w_j(\boldsymbol{x}) = \frac{\partial \mathcal{L}_j(m)}{\partial m}$$
$$= -\frac{2}{m^3(\boldsymbol{x})}\sum_{s=1}^{N_s}\int_0^T q_{j,s}(\boldsymbol{x}, T-t)\frac{\partial^2 u_{j,s}(\boldsymbol{x}, t)}{\partial t^2}dt. \tag{7}$$

The wavefield $u_{j,s}(\boldsymbol{x}, t)$ is now the receiver-specific forward field that is generated using the velocity model m and the current shot s as input signal. Different from the central case, the adjoint wavefield $q_{j,s}(\boldsymbol{x}, t)$ is generated by injecting only the local receiver-specific residual $\boldsymbol{d}_{\text{obs},j} - \boldsymbol{d}_{\text{syn},j}(m)$ at the receiver position $\boldsymbol{x}_j$ per shot s. It therefore satisfies the following wave equation:

$$\left(\frac{1}{m^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \Delta\right)q_{j,s}(\boldsymbol{x}, t) = (\boldsymbol{d}_{\text{obs},j} - \boldsymbol{d}_{\text{syn},j}(m))\,\delta(\boldsymbol{x} - \boldsymbol{x}_j). \tag{8}$$

By comparing the local residual field (8) to the global one in (4) we can deduce that $q_s(\boldsymbol{x}, t) = \sum_{j\in\mathcal{J}} q_{j,s}(\boldsymbol{x}, t)$. Hence, we obtain the important result that the gradient $w(\boldsymbol{x})$ of the global cost $\mathcal{L}(m)$ is equivalent to the sum over all local gradients in the network:

$$w(\boldsymbol{x}) = -\frac{2}{m^3(\boldsymbol{x})}\sum_{j\in\mathcal{J}}\sum_{s=1}^{N_s}\int_0^T q_{j,s}(\boldsymbol{x}, t-T)\frac{\partial^2 u_{j,s}(\boldsymbol{x}, t)}{\partial t^2}dt$$
$$= \sum_{j\in\mathcal{J}} w_j(\boldsymbol{x}) \tag{9}$$

To compute the global gradient in a distributed fashion we employ the adapt-then-combine (ATC) strategy which is a distributed optimization method [10]. The ATC consists of a local model update and a local model fusion step where both steps incorporate a weighted average of neighboring gradients and models. To apply the ATC strategy to the FWI, we introduce a local velocity model $\boldsymbol{m}_j^{[k]}$ and an intermediate model $\widetilde{\boldsymbol{m}}_j^{[k]}$ per receiver j and iteration k. We can then derive the ATC update equations at receiver j as follows:

$$\widetilde{\boldsymbol{m}}_j^{[k+1]} = \boldsymbol{m}_j^{[k]} + \alpha^{[k]}\sum_{i\in\mathcal{N}_j}c_{ij}\boldsymbol{w}_i^{[k]} \tag{10a}$$

$$\boldsymbol{m}_j^{[k+1]} = \sum_{i\in\mathcal{N}_j}a_{ij}\widetilde{\boldsymbol{m}}_i^{[k+1]} \tag{10b}$$
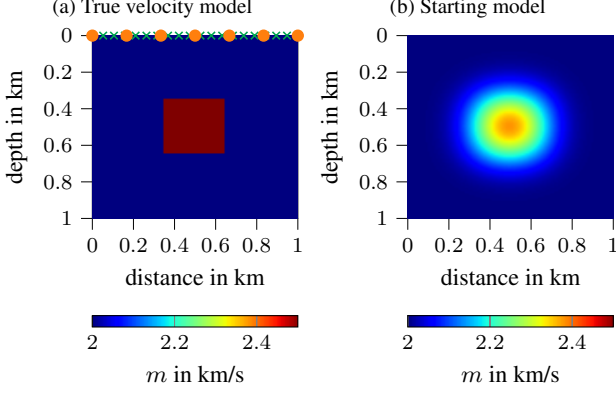
Fig. 1: (a) True velocity model with receivers (✕) and shots (●) placed on surface and (b) smoothed version as starting model.

The coefficients $a_{ij}, c_{ij} \in \mathbb{R}$ determine how data from neighbor i is weighted for the averaging operation at receiver j. They need to satisfy the conditions [17, Eq. (23)] to guarantee convergence towards the global model. Here, we specifically choose these coefficients to

$$a_{ij} = c_{ij} = \begin{cases} 1/|\mathcal{N}_j| & \text{if } i \in \mathcal{N}_j, \\ 0 & \text{else.} \end{cases} \quad (11)$$

An overview of other possible coefficient weights can be found in [18, Section 8]. From (10) we observe that two averaging operations are performed per iteration k: one over the gradients $\boldsymbol{w}_i^{[k]}$ and another one over the models $\widetilde{\boldsymbol{m}}_i^{[k+1]}$. Applying these operations iteratively eventually enable each receiver j to obtain an estimate of the sum of all local gradients, i.e., the global gradient in (9). Hence, each receiver-specific model $\boldsymbol{m}_j^{[k]}$ converges towards the global model $\boldsymbol{m}$ for strongly connected networks [10]. In each iteration k the gradients $\boldsymbol{w}_i^{[k]}$ and models $\widetilde{\boldsymbol{m}}_i^{[k+1]}$ need to be exchanged among neighboring receivers. Algorithm 1 summarizes the proposed adapt-then-combine full waveform inversion (ATC-FWI). To simplify notation and stick to the introduced formalism we use both continuous and discrete variables here. Since ATC-FWI relies on the acoustic wave equation (2), it can be applied to any other imaging application that uses the same wave equation.

5. NUMERICAL EVALUATION

In the following, we verify our proposed method in numerical simulations. We consider a subsurface of an area of $1\,\text{km} \times 1\,\text{km}$ with a background velocity of $m_{\text{bg}} = 2\,\text{km/s}$ that contains a rectangular anomaly in its center with a velocity of $m_{\text{an}} = 2.5\,\text{km/s}$, cf. Figure 1a. We employ a network of $J = 20$ receivers placed evenly in a line on the surface. Accordingly, for the network we assume a line topology where each receiver is connected to its next three receivers to its left and right side. Thus, in total each receiver has at most six neighbors and at least three (i.e., for the first and last receiver in the line). For the exploration survey we use $N_s = 7$ shots evenly placed on the surface. The input source is a Ricker wavelet with center frequency $f = 10\,\text{Hz}$ and $0°$ phase shift. The total observation time is $T = 1\,\text{s}$. To solve the wave equation numerically we use the Devito package [19] which is a symbolic finite-difference framework for partial differential equations. For the discretization by finite-differences we set each spatial cell to a size of $10\,\text{m} \times 10\,\text{m}$ and use absorbing boundaries with 20 grid cells.


Algorithm 1 Adapt-then-combine FWI (ATC-FWI)

1: Initialize starting model $m^{[0]}$
2: **for** FWI iteration $k \leftarrow 0, N_{\text{FWI}} - 1$ **do**
3:    **for** receiver $j \leftarrow 1, J$ **do**   ▷ Local gradient computation
4:       $w_j^{[k]} = 0$
5:       **for** shot $s \leftarrow 1, N_s$ **do**
6:          Solve forward model with $m_j^{[k]} \rightarrow u_{j,s}(\boldsymbol{x}, t), \boldsymbol{d}_{\text{syn},j}$
7:          Compute residual wavefield $q_{j,s}(\boldsymbol{x}, T - t)$
8:          Accumulate gradients:
9:          $w_j^{[k]} \leftarrow w_j^{[k]} - \int_0^T q_{j,s}(\boldsymbol{x}, T - t)\frac{\partial^2 u_{j,s}(\boldsymbol{x},t)}{\partial t^2} dt$
10:       **end for**
11:    **end for**
12:    **for** receiver $j \leftarrow 1, J$ **do**   ▷ Adapt-then-combine
13:       $\widetilde{\boldsymbol{m}}_j^{[k+1]} = \boldsymbol{m}_j^{[k]} + \alpha^{[k]} \sum_{i \in \mathcal{N}_j} c_{ij} \boldsymbol{w}_i^{[k]}$
14:       $\boldsymbol{m}_j^{[k+1]} = \sum_{i \in \mathcal{N}_j} a_{ij} \widetilde{\boldsymbol{m}}_i^{[k+1]}$
15:    **end for**
16: **end for**
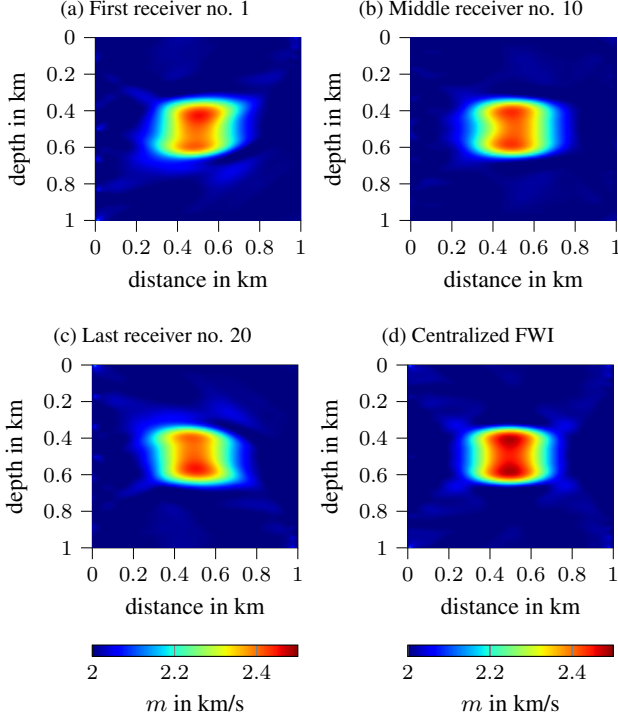17: **return** Local model $\boldsymbol{m}_j^{[N_{\text{FWI}}]}, \forall j \in \mathcal{J}$
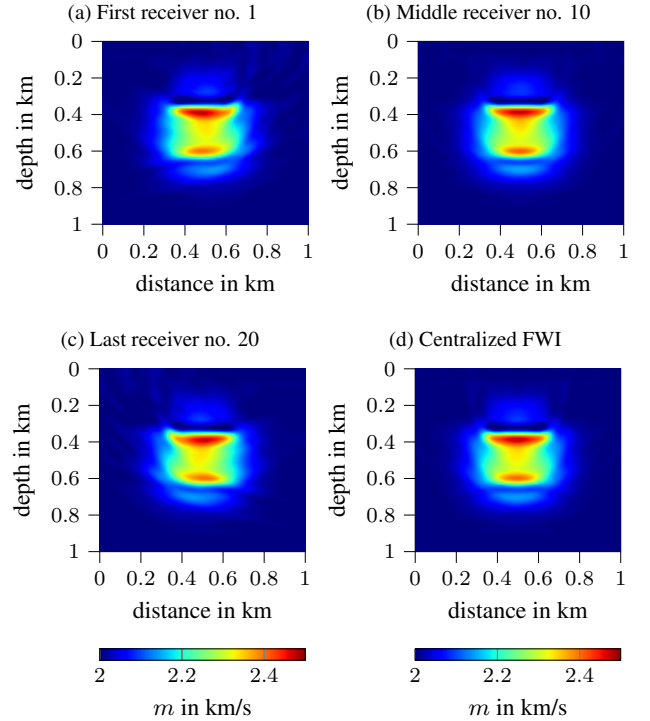
Fig. 2: Estimated velocity models using ATC-FWI and centralized FWI after 20 iterations.

We apply both the centralized FWI and our proposed ATC-FWI to the reconstruction of the above described subsurface. To stabilize convergence behavior we choose the step size to $\alpha^{[k]} = 0.01/\max\{\boldsymbol{w}_j^{[k]}\}$ for ATC-FWI and $\alpha^{[k]} = 0.01/\max\{\boldsymbol{w}^{[k]}\}$ for centralized FWI, respectively. For both algorithms we use a smoothed version of the true velocity model as starting velocity model $\boldsymbol{m}^{[0]}$ by applying a Gaussian filter with standard deviation $\sigma = 10$ to it, cf. Figure 1b.

Figure 2 illustrates the estimated velocity models for ATC-FWI

(a) First receiver no. 1   (b) Middle receiver no. 10

(c) Last receiver no. 20   (d) Centralized FWI

m in km/s m in km/s

Fig. 3: Estimated velocity models using ATC-FWI and centralized FWI after 20 iterations for the crosshole constellation.

(a) Surface constellation.   (b) Crosshole constellation.

Fig. 4: Velocity profile at distance $x = 0.5\,\mathrm{km}$ for receiver no. 1 in (a) surface and (b) crosshole constellation.

Fig. 5: Cost function for (a) surface and (b) crosshole constellation.

after $N_{\mathrm{FWI}} = 20$ iterations at three receivers. We can observe that all receivers obtain a velocity model close or equal to the one obtained by the centralized FWI. Receiver 1 and 20 obtain a model that is slightly different from centralized FWI. This is due to the fact that these receivers have the lowest number of neighbors in the network and are therefore not able to achieve the exact global model within the limited amount of $N_{\mathrm{FWI}} = 20$ iterations. On the contrary, we observe that the middle receiver 10 achieves the central model since this receiver has the maximum number of neighbors and can therefore fuse all gradients and models in the network over the iterations. In another experiment, we change the source/receiver positions from a surface to a crosshole constellation. The sources are now uniformly placed at a distance $x = 0\,\mathrm{km}$ over the complete depth while receivers are placed uniformly at $x = 980\,\mathrm{km}$ over the complete depth. All other parameters as well as the starting model are kept as in the surface constellation. Crosshole constellations usually lead to better imaging results since seismic rays are shot from one side through the object and received on the opposite side [2]. The resulting velocity models can be seen in Figure 3. Also here ATC-FWI obtains velocity models close to the central model where the result of receiver 10 is closest to the central one. Compared to the surface constellation the rectangular anomaly in the center part is more accurately recovered.

In Figure 4a, we show the velocity profile over the depth at a distance of $x = 0.5\,\mathrm{km}$ after $N_{\mathrm{FWI}} = 20$ FWI iterations for the first receiver. Also here we can observe that the estimated velocity profile of ATC-FWI is close to the central profile especially for the surface constellation. One can clearly see how FWI sharpens the edges where a high velocity gradient is present compared to the starting model that has a smooth change in velocity between background
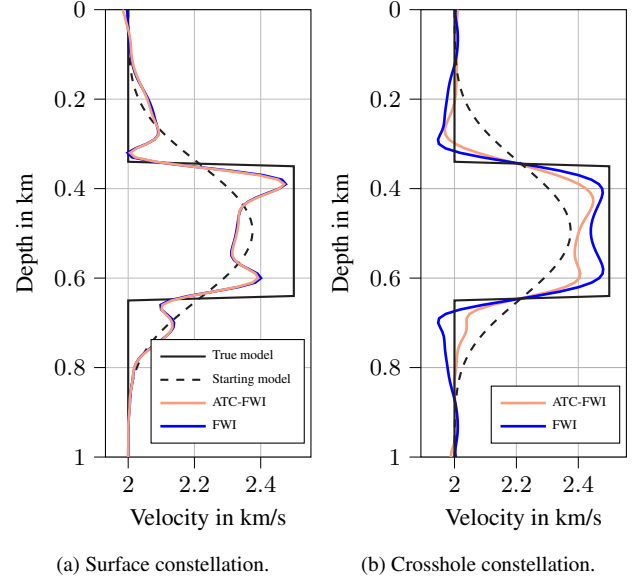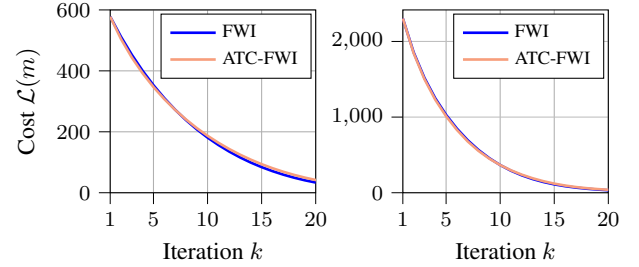
and anomaly. In the crosshole experiment the center region of the anomaly is more accurately reconstructed compared to the surface constellation. However, a larger gap between the models of ATC-FWI and centralized FWI can be observed. Figure 5 shows the cost function $\mathcal{L}(m)$ over the FWI iterations for the surface and crosshole constellation. For ATC-FWI the cost is summed up over all receivers to be comparable to the centralized case. In both cases the cost of ATC-FWI is nearly identical to the central case.

6. CONCLUSION AND FUTURE WORK

In this paper, we showed that the global gradient in FWI is equivalent to the sum of local gradients. Based on this result, we proposed the use of the adapt-then-combine method to estimate the global gradient locally at each receiver and to enable a distributed implementation of the FWI. For numerical evaluations we used synthetic velocity models and showed that the ATC-FWI achieves similar performance as the original, centralized FWI. For future work, we intend to apply our proposed method to real seismic data and to extend the method to 3D-exploration tasks.

7. REFERENCES

[1] Albert Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984.

[2] Jean Virieux and Stephane Operto, "An overview of full-waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, 2009.

[3] Romain Brossier, Stéphane Operto, and Jean Virieux, "Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion," *Geophysics*, vol. 74, no. 6, 2009.

[4] Yunyue Elita Li and Laurent Demanet, "Full waveform inversion with extrapolated low frequency data," *Offshore Technology Conference Asia 2016, OTCA 2016*, , no. 2008, pp. 1929–1935, 2016.

[5] Eran Treister and Eldad Haber, "Full Waveform Inversion Guided by Travel Time Tomography," *SIAM Journal on Scientific Computing*, vol. 39, no. 5, pp. S587–S609, 2017.

[6] Lluís Guasch, Oscar Calderón Agudo, Meng Xing Tang, Parashkev Nachev, and Michael Warner, "Full-waveform inversion imaging of the human brain," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–12, 2020.

[7] Ulas Taskin and Koen W. A. van Dongen, "3-D redatuming for breast ultrasound," in *SPIE Medical Imaging*, 2020, number March, p. 199.

[8] Emanuel Staudinger, Dmitriy Shutin, Christoph Manß, Alberto Viseras Ruiz, and Siwei Zhang, "Swarm technologies for future space exploration missions," in *International Symposium on Artificial Intelligence, Robotics and Automation in Space*, 2018.

[9] Soummya Kar and Jose M.F. Moura, "Consensus + innovations distributed inference over networks: cooperation and sensing in networked systems," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 99–109, may 2013.

[10] Ali H. Sayed, Shengyuan Tu, Jianshu Chen, Xiaochuan Zhao, and Zaid J Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Processing Magazine*, pp. 155–171, 2013.

[11] A. Kadu and R. Kumar, "Decentralized full-waveform inversion," in *80th EAGE Conference and Exhibition 2018: Opportunities Presented by the Energy Transition*, 2018, number January.

[12] Toktam Zand, Hamid R. Siahkoohi, Alison Malcolm, Ali Gholami, and Alan Richardson, "Consensus optimization of total variation–based reverse time migration," *Computational Geosciences*, pp. 1393–1407, 2020.

[13] Florian Xaver, Gerald Matz, Peter Gerstoft, and Christoph Mecklenbräuker, "Localization of acoustic sources using a decentralized particle filter," *Eurasip Journal on Wireless Communications and Networking*, vol. 2011, no. 1, pp. 1–14, 2011.

[14] Fangyu Li, Yan Qin, and Wenzhan Song, "Waveform Inversion-Assisted Distributed Reverse Time Migration for Microseismic Location," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1327–1332, 2019.

[15] Rene Edouard Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 2006.

[16] Jon F. Claerbout, "Toward a unified theory of reflector mapping," *Geophysics*, vol. 36, no. 3, pp. 467–481, 1971.

[17] Frederico S. Cattivelli and Ali H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.

[18] Ali H. Sayed, "Chapter 9 - diffusion adaptation over networks," in *Academic Press Library in Signal Processing: Volume 3*, vol. 3, pp. 323 – 453. Elsevier, 2014.

[19] 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*, vol. 12, no. 3, pp. 1165–1187, 2019.