Joint Distributed Traveltime and Full Waveform Tomography for Enhanced Subsurface Imaging in Seismic Networks

Ban-Sok Shin, *Member, IEEE,* and Dmitriy Shutin, *Senior Member, IEEE*

***Abstract*—Imaging the subsurface of extra-terrestrial bodies and planets has gained significant attention in recent space exploration missions. Multi-agent systems that autonomously perform subsurface imaging have been proposed. There, agents collaborate to image the subsurface by leveraging their wireless connections, enabling each agent to obtain an estimate of the subsurface image. However, traditional subsurface imaging techniques rely on a single entity for data collection and inversion, making them centralized schemes that limit direct availability of subsurface images at the agents. In this article, we propose the joint use of distributed seismic imaging techniques based on traveltime tomography and full waveform inversion, namely, the distributed traveltime tomography (D-TOMO) and the adapt-then-combine full waveform inversion (ATC-FWI). Combined in a sequential manner these techniques allow each agent to infer high-resolution subsurface images by exchanging data with their neighboring agents starting from a simple initial subsurface model. Unlike existing decentralized seismic imaging methods, our proposed scheme is fully distributed and provides flexibility without the need of anchor nodes or a full mesh topology. We demonstrate that ATC-FWI can recover high-frequency components based on a low-resolution subsurface image provided by D-TOMO in the initial stage. To assess the imaging performance, we employ a synthetic model, the SEG/EAGE salt model as well as real data from field measurements conducted over a tunnel.**

***Index Terms*—Distributed imaging, traveltime tomography, full waveform inversion, seismic imaging, seismic exploration, multi-agent networks, inverse problems**

I. Introduction

A. Motivation

GEOPHYSICAL imaging for extra-terrestrial applications has become increasingly relevant in the last decade. Recent space missions dedicate substantial resources to gain more insight into the global composition and geophysical structure of e.g. Mars or Moon [1], [2], [3]. Future space missions will also focus on providing more detailed insights to the near-surface of such planets to reveal anomalies such as caves or lava tubes [4]. Such anomalies are of high interest as they could be used as safe habitats for humans and space equipment that might be harmed on the surface. To discover such anomalies concepts of seismic exploration surveys conducted by a multi-agent network have been proposed in the past [5], [6]. Here, the vision is to use multiple robotic agents connected to each other over wireless links to autonomously image a subsurface in a cooperative fashion. This concept is illustrated in Fig. 1. Based on the reconstructed subsurface image the agents shall then optimize their sampling positions to enhance certain features in the image or the whole image itself. However, in traditional seismic imaging all measured data are collected at a central entity that performs the inversion. Hence, the reconstructed image is not available locally at the agents but at the central entity only. To address this issue, distributed subsurface imaging is required that provides each agent with an estimate of the subsurface.

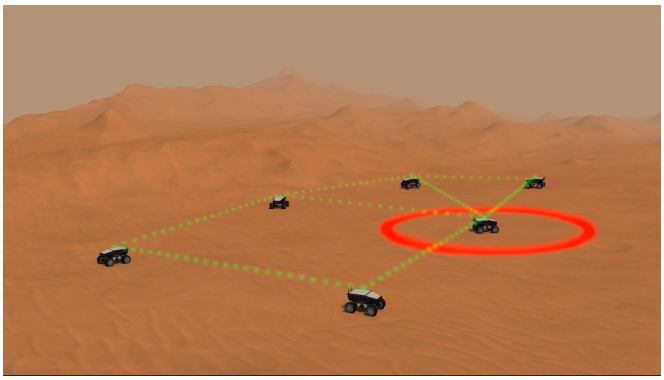


Fig. 1: Conceptual illustration of a multi-agent system that autonomously explores a subsurface using seismic imaging. The agents are connected to each other over wireless links. Each agent is equipped with a geophone to measure seismic data. One agent acts as source by injecting seismic waves into the subsurface as indicated by the red circle.

In relation to seismic networks and decentralized processing of seismic data, several works have been proposed recently. In [7] a distributed traveltime tomography for seismic networks is proposed. The authors employ a Bayesian algebraic reconstruction technique to obtain tomographic images of the subsurface in large-scale networks. The images can be obtained within the seismic network. However, the network contains a sink node that still needs to gather data from all nodes, performs an average and distributes it back to all nodes. Hence, it is not a completely distributed processing method since a central entity is needed. The methods [8], [9] rely on similar approaches for ambient noise tomography with possibly multiple sink nodes that collect data only from nodes directly connected to them. The sink nodes need to be connected in a full mesh topology to obtain data over the complete area of interest which poses a strong restriction.

B. Shin and D. Shutin are with the Institute of Communications and Navigation, German Aerospace Center, 82234 Wessling, Germany, e-mails: ban-sok.shin@dlr.de, dmitriy.shutin@dlr.de.

This research is partly funded by the German Research Foundation (DFG) under grant SH 1975/1-1 and SH 1743/3-1.

Manuscript received April 19, 2005; revised August 26, 2015.

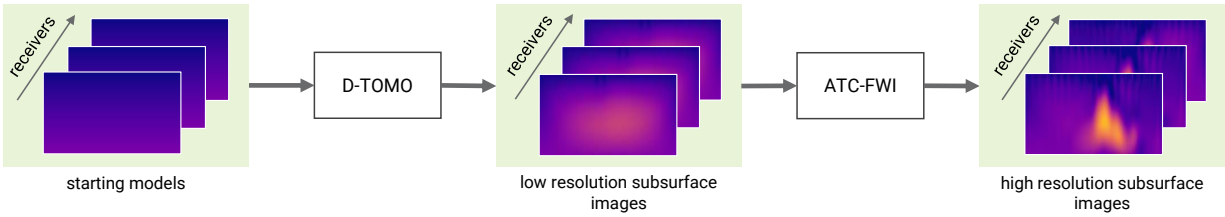


Fig. 2: Overview of the proposed joint distributed imaging scheme: Each receiver is initialized with the same starting model, e.g., a model of gradually increasing velocity. D-TOMO obtains low-resolution subsurface images in a distributed way at each receiver. These are then used as initial models for the ATC-FWI which enhances the resolution substantially.

Unlike these methods we proposed fully distributed imaging schemes based on traveltime tomography (TT) and full waveform inversion (FWI). Specifically, we developed the distributed traveltime tomography (D-TOMO) [12] and the adapt-then-combine full waveform inversion (ATC-FWI) [13]. These schemes enable each agent to obtain a subsurface image that resembles the traditional, centralized imaging result. Our proposed algorithms offer a higher flexibility compared to the aforementioned distributed imaging schemes in the literature. For instance, we do not require any central entity such as anchor nodes. Rather each agent in the network has the same functionality and obtains a subsurface image locally via cooperation with neighboring agents. Furthermore, our network topology can be flexibly changed and does not require a full mesh topology. This is highly relevant for an autonomous exploration by rovers where the network topology is changed due to movement of the rovers.

B. Main Contributions

In this article, we propose and investigate a joint distributed subsurface imaging scheme that combines both D-TOMO and ATC-FWI. D-TOMO obtains subsurface images of low spatial resolution at each agent/node in a distributed fashion. We propose to use these image as initial subsurface models for the ATC-FWI which aims at enhancing their spatial resolution in a distributed way enabling each node to obtain high-resolution subsurface images locally. Fig. 2 illustrates our proposed joint scheme. Our contributions can be summarized as follows:

1) In literature it is common to initialize the FWI with a blurred version of the ground truth image [14], [15]. This was also done in our previous study for the ATC-FWI [13]. In contrast to that, here we investigate the effect of obtaining an initial subsurface image for ATC-FWI by D-TOMO. These initial images can differ over the nodes due to the distributed computing nature of D-TOMO. We also demonstrate that only a low number of iterations for D-TOMO is sufficient to still obtain high-resolution images by ATC-FWI. In the extreme case, even only one iteration by D-TOMO is suitable.
2) In contrast to our previous work, we investigate both D-TOMO and ATC-FWI using seismic benchmark data and more importantly, real data obtained in a seismic field experiment. The benchmark data consists of the SEG/EAGE salt model [16] that is widely utilized as a benchmark velocity model in seismic imaging. We demonstrate that our proposed joint scheme allows for high-resolution distributed imaging starting from a simple initial velocity model at each node. As real dataset we employ seismic field data that we recorded over a highway tunnel using hammer strikes as an active source. In particular, we show that D-TOMO is able to image the tunnel anomaly in the subsurface and that the obtained images match available ground truth data on the tunnel.
3) As a further contribution, we provide a rigorous derivation of the gradient in adjoint-state-based TT that is used in D-TOMO. To this end, we rely on functional derivatives. To the best of our knowledge there is no accurate gradient derivation for this method that uses functional derivatives in the literature.

II. Brief Review of Seismic Imaging Methods

We start by briefly introducing the seismic imaging methods used in our proposed algorithms. Both TT and FWI consider the minimization of a least-squares cost between observed and synthesized data based on an estimate of the subsurface model. To obtain synthesized data, in both methods a partial differential equation (PDE) needs to be solved. This results in a highly nonlinear inverse problem with respect to (w.r.t.) the subsurface model with the cost function having multiple local minima. Both TT and FWI apply iterative optimization methods to approach a point close to the global minimum by computing a gradient and updating the subsurface estimate accordingly. To solve the respective PDE, numerical methods are required. Here, we rely on finite differences and discretize the computational domain in a regular grid of $N_x \times N_z$ cells for the x and z-axis, respectively. We start our description in the continuous domain and then transfer results to the discretized domain for the model updates.

A. Traveltime Tomography (TT)

TT images geophysical subsurface structures based on measured traveltimes between a source and multiple receivers or geophones. It relies on first-arrival traveltimes in the measured seismograms, i.e., the time of the event of an incoming wave at a specific receiver. In mathematical terms, TT considers the minimization of the squared residual between measured and synthesized traveltimes at the receiver positions for N_R receivers and N_S sources with respect to a spatial velocity function $m(\boldsymbol{x})$. The scalar function $m(\boldsymbol{x})$ represents the P-wave velocity distribution over the spatial coordinate $\boldsymbol{x} =$

$(x, z) \in \mathbb{R}^2$ in the subsurface domain $\Omega \subset \mathbb{R}^2$. The corresponding cost function to be minimized is given by [17]

$$\min_m \mathcal{J}_{\mathrm{TT}}(m) = \frac{1}{2}\sum_{s=1}^{N_\mathrm{S}} \int_{\partial\Omega} |T_s(m, \boldsymbol{x}) - T_s^{\mathrm{obs}}(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x}, \quad (1)$$

where T_s^{obs} and T_s are, respectively, the measured and synthesized traveltimes for source s. $\partial\Omega$ is the boundary of the subsurface domain along which the receivers are placed at positions $\boldsymbol{x}_r \in \mathbb{R}^2$. To keep notation uncluttered we omit the dependence of m on $\boldsymbol{x}$ here. The measured traveltimes T_s^{obs} are picked either automatically or by hand from the seismogram at each receiver r. In problem (1) the traveltimes $T_s(m, \boldsymbol{x})$ need to be synthesized based on the current velocity model estimate $m(\boldsymbol{x})$ to obtain the residuals. This can be achieved by solving the *eikonal equation* given as

$$|\nabla T_s(\boldsymbol{x})|^2 = \frac{1}{m(\boldsymbol{x})^2}, \quad \text{s.t.} \quad T_s(\boldsymbol{x}_s) = 0, \boldsymbol{x} \in \Omega \quad (2)$$

w.r.t. the traveltime function $T_s(\boldsymbol{x})$ for a shot s. The associated constraint demands that the traveltime is zero at source location $\boldsymbol{x}_s \in \mathbb{R}^2$. Solving (2) with respect to $T_s(\boldsymbol{x})$ gives a map of traveltimes over $\boldsymbol{x}$ which is then sampled at the receiver locations to give $T_s(m, \boldsymbol{x}_r)$. Note that solving the inverse problem in this setting requires solving (2) multiple times. To this end, several numerical methods exist, e.g. the fast marching method [18], fast sweeping method [19] or the fast iterative method [20].

Since TT utilizes an iterative, gradient-based minimization scheme to estimate $m(\boldsymbol{x})$ the gradient of $\mathcal{J}_{\mathrm{TT}}(m)$ is required. To calculate such gradient ray tracing has been applied in the past to linearize the traveltime around the current velocity model m [21], [22]. An alternative is the *adjoint-state method*. This method is a general technique to solve optimization problems that are constrained by a PDE [23]. Here, an additional variable, the adjoint-state $\lambda_s(\boldsymbol{x})$, is introduced per shot s to account for the PDE constraint. Applying the adjoint-state method to problem (1), we obtain the following gradient:

$$\frac{\mathrm{d}\mathcal{J}_{\mathrm{TT}}(m)}{\mathrm{d}m} = -\sum_{s=1}^{N_\mathrm{S}} \frac{\lambda_s(\boldsymbol{x})}{m(\boldsymbol{x})^3}, \quad (3)$$

where $\mathrm{d}/\mathrm{d}m$ denotes the total derivative w.r.t. m. The adjoint-state $\lambda_s(\boldsymbol{x})$ for shot s is computed by solving the following (adjoint) PDE:

$$\nabla \cdot (\lambda_s(\boldsymbol{x}) \nabla T_s(\boldsymbol{x})) = 0, \; \boldsymbol{x} \in \Omega \quad (4a)$$

$$\lambda_s(\boldsymbol{x}_r) \left(\nabla T_s(\boldsymbol{x}_r) \cdot \boldsymbol{n}\right) = T_s(\boldsymbol{x}_r) - T_s^{\mathrm{obs}}(\boldsymbol{x}_r), \; \boldsymbol{x}_r \in \partial\Omega \quad (4b)$$

By $(\cdot)$ we denote the dot product between two vector fields. A detailed derivation of the gradient in (3) and the adjoint-state equations in (4) is given in Appendix A and B. To compute the adjoint-state, we initialize $\lambda_s(\boldsymbol{x})$ by the residuals $T_s(\boldsymbol{x}_r) - T_s^{\mathrm{obs}}(\boldsymbol{x}_r)$ following the boundary condition (4b) where $\boldsymbol{n}$ is the unit vector normal to the boundary $\partial\Omega$. Then (4a) is used to solve for the adjoint-state $\lambda_s(\boldsymbol{x})$ in the inner domain Ω. To this end, we employ the fast sweeping method, cf. [24], [17]. With the adjoint-state $\lambda_s(\boldsymbol{x})$ the gradient is computed following (3). However, authors in [24] describe that such gradient computation might lead to instabilities in an iterative minimization procedure. Thus in [24, Eq. (2.17)], the authors proposed the following gradient smoothing:

$$(I - \nu\Delta) m_\Delta(\boldsymbol{x}) = -\sum_{s=1}^{N_\mathrm{S}} \frac{\lambda_s(\boldsymbol{x})}{m(\boldsymbol{x})^3} \quad (5)$$

with $m_\Delta(\boldsymbol{x})$ being the smoothed gradient, I the identity operator and Δ the Laplace operator. Parameter $\nu \geq 0$ controls the smoothness of $m_\Delta(\boldsymbol{x})$. As described, we use finite differences to discretize domain Ω. This results in vector representations for the velocity model $m(\boldsymbol{x})$, the adjoint-states $\lambda_s(\boldsymbol{x})$ and the gradient $m_\Delta(\boldsymbol{x})$ that we denote via $\boldsymbol{m} \in \mathbb{R}^{N_x N_z}, \boldsymbol{m}_\Delta \in \mathbb{R}^{N_x N_z}$ and $\boldsymbol{\lambda}_s \in \mathbb{R}^{N_x N_z}$, respectively. To minimize the cost $\mathcal{J}_{\mathrm{TT}}(m)$ iteratively gradient descent can be used as

$$\boldsymbol{m}^{[k+1]} = \boldsymbol{m}^{[k]} - \alpha_{\mathrm{TT}}^{[k]} \boldsymbol{m}_\Delta. \quad (6)$$

The parameter $\alpha_{\mathrm{TT}}^{[k]} > 0$ is a step size parameter that needs to be chosen appropriately to guarantee convergence. Other optimization schemes such as nonlinear conjugate gradient or the L-BFGS can be applied here to accelerate convergence.

B. Full Waveform Inversion (FWI)

Beside TT another subsurface imaging technique is the full waveform inversion. Different from TT it uses the whole seismic trace that is measured by the receivers. While TT uses the eikonal equation to generate traveltimes, FWI employs the acoustic or elastic wave equation to synthesize seismic traces [15]. The general objective in FWI is similar to TT: minimize a residual error between observed and synthesized seismic data with respect to a subsurface model parameter m such as the P-wave velocity, S-wave velocity, etc.

In its basic form, FWI considers a least-squares cost between measured seismic data $d_{s,r}^{\mathrm{obs}}$ and synthesized seismic data $d_{s,r}^{\mathrm{syn}}$ at receiver r for source s over a measurement time τ [15]. More formally,

$$\min_m \mathcal{J}_{\mathrm{FWI}}(m) = \frac{1}{2}\sum_{s=1}^{N_\mathrm{S}}\sum_{r=1}^{N_\mathrm{R}} \int_0^\tau \left(d_{s,r}^{\mathrm{syn}}(t, m) - d_{s,r}^{\mathrm{obs}}(t)\right)^2 \mathrm{d}t \quad (7)$$

where $d_{s,r}^{\mathrm{syn}}(t, m) = S_{s,r} u_s(\boldsymbol{x}, t, m)$ is obtained by sampling the wavefield $u_s(\boldsymbol{x}, t, m)$ via the operator $S_{s,r}$ at the receiver positions $\boldsymbol{x}_r$. For a P-wave velocity model $m(\boldsymbol{x})$, the wavefield $u_s(\boldsymbol{x}, t, m)$ is computed by solving the acoustic wave equation

$$\frac{1}{m(\boldsymbol{x})^2}\frac{\partial^2 u_s(\boldsymbol{x}, t)}{\partial t^2} - \Delta u_s(\boldsymbol{x}, t) = f_s(\boldsymbol{x}, t) \quad (8)$$

with initial conditions

$$u_s(\boldsymbol{x}, 0) = 0, \quad \frac{\partial u_s(\boldsymbol{x}, 0)}{\partial t} = 0. \quad (9)$$

Here, $f_s(\boldsymbol{x}, t)$ is the seismic source and $u_s(\boldsymbol{x}, t)$ is the wavefield at position $\boldsymbol{x}$ and time t that describes the wave propagation through the subsurface. Note that (8) describes the propagation of P-waves only. Thus, conversion into other wave types such as S-waves at subsurface interfaces is not represented. To solve (8) numerically, a variety of solvers can

be used based on finite differences, finite elements, spectral elements, etc. [25].

To minimize (7) iteratively we need the gradient of $\mathcal{J}_{\mathrm{FWI}}(m)$ with respect to m. Again the adjoint-state method is used for this purpose. Following the adjoint-state method [23], the gradient is obtained via

$$\frac{\mathrm{d}\mathcal{L}_{\mathrm{FWI}}(m)}{\mathrm{d}m} = -\frac{2}{m(\boldsymbol{x})^3}\sum_{s=1}^{N_{\mathrm{S}}}\int_0^{\tau} q_s(\boldsymbol{x},\tau-t)\frac{\partial^2 u_s(\boldsymbol{x},t)}{\partial t^2}\mathrm{d}t. \tag{10}$$

A derivation of the gradient is given in Appendix C. The variable $q_s(\boldsymbol{x},\tau-t)$ is the adjoint wavefield that is obtained by solving another wave equation given as

$$\left(\frac{1}{m(\boldsymbol{x})^2}\frac{\partial^2}{\partial t^2}-\Delta\right)q_s(\boldsymbol{x},t) = \sum_{r=1}^{N_{\mathrm{R}}} S_{s,r}^{\mathsf{T}}(d_{s,r}^{\mathrm{syn}}(\tau-t) - d_{s,r}^{\mathrm{obs}}(\tau-t)) \tag{11}$$

with initial conditions $q_s(0)=0, \partial q_s(0)/\partial t = 0$. To solve (11), the data residuals $d_{s,r}^{\mathrm{syn}}(\tau-t) - d_{s,r}^{\mathrm{obs}}(\tau-t)$ are first reversed in time and placed at the receiver positions as source terms for the wave equation. Then (11) is solved w.r.t. $q_s(\boldsymbol{x},t)$. To calculate the gradient following (10), the adjoint wavefield $q_s(\boldsymbol{x},\tau-t)$ is again reversed back in time and correlated with the second derivative of the forward wavefield $u_s(\boldsymbol{x},t)$. Such correlation is known as the *imaging condition* or *migration* [26]. Afterwards gradient descent is used to iteratively minimize the cost $\mathcal{J}_{\mathrm{FWI}}(m)$. After discretization of the spatial domain Ω we stack all velocity values and all gradient values into vectors $\boldsymbol{m}^{[k]} \in \mathbb{R}^{N_x N_z}$ and $\boldsymbol{m}_{\Delta}^{[k]} \in \mathbb{R}^{N_x N_z}$, respectively, at iteration k. The current velocity model is then updated using

$$\boldsymbol{m}^{[k+1]} = \boldsymbol{m}^{[k]} - \alpha_{\mathrm{FWI}}^{[k]}\boldsymbol{m}_{\Delta}^{[k]}, \tag{12}$$

where $\alpha_{\mathrm{FWI}}^{[k]} > 0$ is a step size that needs proper selection either manually or by a line search method.

III. Joint Distributed Subsurface Imaging

The imaging methods presented so far are centralized methods, i.e., all measurement data is assumed to be available at a single entity that performs the inversion. In the following, we briefly review both D-TOMO and ATC-FWI proposed in our earlier works that allow each receiver in a seismic network to obtain a subsurface image locally via cooperation with other receivers and without dependence on a central processing entity. Afterwards, we discuss their joint operation. We start by introducing a model for the seismic network.

A. Seismic Network Model

We consider a seismic network of N_{R} receivers/geophones placed in a line array for a 2D subsurface imaging task. The topology of the seismic network is described via a graph $\mathcal{G} = \{\mathcal{R},\mathcal{E}\}$ with a set of nodes $\mathcal{J} = \{1,2,\ldots,N_{\mathrm{R}}\}$ and a set of edges $\mathcal{E} = \{(r,\ell)|r,\ell \in \mathcal{R}, r \neq \ell\}$ for wireless connections among the receivers. We assume that the graph $\mathcal{G}$ is undirected and strongly connected [27], i.e., each receiver can be reached by any other receiver in the network over multiple hops. Furthermore, we define a neighborhood set

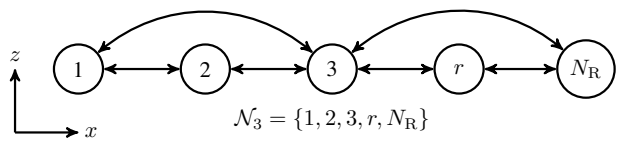


Fig. 3: Example of a line array network topology with neighborhood set $\mathcal{N}_3$ of receiver 3.

$\mathcal{N}_r$ for each receiver j consisting of those receivers that are directly connected to receiver r and of the receiver r itself. Each neighbor ℓ contained in $\mathcal{N}_r$ can exchange data with receiver r and vice versa. Moreover, we assume that all receiver positions $\{\boldsymbol{x}_r\}_{r=1}^{N_{\mathrm{R}}}$ are known to each receiver r. An example of the network topology is shown in Fig. 3. For subsurface imaging, we assume N_{S} arbitrary shot locations on the surface and a total measurement time of τ per shot $s = \{1,2,\ldots,N_{\mathrm{S}}\}$. Each receiver r has a fixed Cartesian coordinate denoted by $\boldsymbol{x}_r = (x_r, z_r)$ with the x- and z-coordinate.

B. Distributed Traveltime Tomography (D-TOMO)

In the following, we briefly review the D-TOMO as a distributed traveltime tomography method. For further details the reader is referred to [12].

The key observation that enables each receiver to obtain an image locally is that the traveltime residuals $\rho_r = T_s(\boldsymbol{x}_r) - T_s^{\mathrm{obs}}(\boldsymbol{x}_r)$ of all receivers $r = 1,\ldots,N_{\mathrm{R}}$ are the necessary data to compute the adjoint-state field $\lambda_s(\boldsymbol{x})$ in (4b) and by that also the gradient $m_{\Delta}(\boldsymbol{x})$ in (5). Thus, if all traveltime residuals $\{\rho_r\}_{r=1}^{N_{\mathrm{R}}}$ are available at each receiver r, the adjoint-state field and the gradient can be computed locally and an update of the subsurface image $m(\boldsymbol{x})$ can be performed by each receiver individually. However, the traveltime residuals are distributed over the receivers, one at each corresponding receiver. To make all traveltime residuals available at each receiver, we proposed to perform a *distributed regression* of these residuals within the seismic network. To this end, we model the distribution of traveltime residuals ρ_r over the receiver positions $\boldsymbol{x}_r$ via a nonlinear function $\rho(\boldsymbol{x})$. The distributed regression will then provide each receiver with an estimate of the function $\rho(\boldsymbol{x})$ and by that with an estimate of all traveltime residuals. To facilitate distributed regression we employ the kernel distributed consensus-based estimation (KDiCE) algorithm proposed in [28]. This algorithm uses kernels to perform a regression of nonlinear functions in a distributed fashion within a network of nodes. The function $\rho(\boldsymbol{x})$ is then modeled via a linear combination of Gaussian kernels with kernel bandwidth $\sigma \in \mathbb{R}$ such that all traveltime residuals $\boldsymbol{\rho} = [\rho_1,\ldots,\rho_{N_{\mathrm{R}}}]^{\mathsf{T}} \in \mathbb{R}^{N_{\mathrm{R}}}$ are approximated by the linear combination $\boldsymbol{G}\boldsymbol{w}_r$. The matrix $\boldsymbol{G} \in \mathbb{R}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$ is a symmetric Gram matrix containing Gaussian kernel evaluations whereas $\boldsymbol{w}_r$ represents the receiver-specific weight vector that combines the Gaussian kernels correspondingly. Then a distributed least-squares problem is formulated w.r.t

the weights $\boldsymbol{w}_r$:

$$\{\boldsymbol{w}_r^\star | r \in \mathcal{R}\} = \underset{\{\boldsymbol{w}_r | r \in \mathcal{R}\}}{\operatorname{argmin}} \sum_{r=1}^{N_\mathrm{R}} ||\boldsymbol{\rho} - \boldsymbol{G}\boldsymbol{w}_r||_2^2 \tag{13a}$$

$$\text{s.t.} \quad \boldsymbol{w}_r = \boldsymbol{w}_\ell, \quad r \in \mathcal{R}, \quad \ell \in \mathcal{N}_r \tag{13b}$$

The additional constraint (13b) is a consensus constraint which introduces convergence to the same weight vector for all receivers over a connected network graph, i.e., $\boldsymbol{w}_1 = \boldsymbol{w}_2 = \ldots = \boldsymbol{w}_{N_\mathrm{R}}$. Problem (13) can be solved in a distributed fashion using the alternating direction method of multipliers, cf. [28], [12]. This results in the following update formulas:

$$\boldsymbol{z}_r^{[n+1]} = \frac{1}{|\mathcal{N}_r|} \sum_{\ell \in \mathcal{N}_r} \left(\boldsymbol{w}_\ell^{[n]} - \varepsilon \boldsymbol{v}_{\ell r}^{[n]} \right), \tag{14a}$$

$$\boldsymbol{v}_{r\ell}^{[n+1]} = \boldsymbol{v}_{r\ell}^{[n]} - \varepsilon^{-1} \left(\boldsymbol{w}_r^{[k]} - \boldsymbol{z}_\ell^{[n+1]} \right), \tag{14b}$$

$$\boldsymbol{w}_r^{[n+1]} = \left[\boldsymbol{g}(\boldsymbol{x}_r)\boldsymbol{g}(\boldsymbol{x}_r)^\mathsf{T} + \varepsilon^{-1}|\mathcal{N}_r|\boldsymbol{I}_{N_\mathrm{R}} \right]^{-1} \times \left\{ \rho_r \boldsymbol{g}(\boldsymbol{x}_r) + \sum_{\ell \in \mathcal{N}_r} \left(\frac{\boldsymbol{z}_\ell^{[n+1]}}{\varepsilon} + \boldsymbol{v}_{r\ell}^{[n+1]} \right) \right\}. \tag{14c}$$

Variable $\boldsymbol{z}_r^{[n]}$ is an auxiliary variable that serves as an intermediate result of the weight vector $\boldsymbol{w}_r$. Note that $\boldsymbol{g}(\boldsymbol{x}_r)\boldsymbol{g}(\boldsymbol{x}_r)^\mathsf{T}$ in (14c) is a rank-1 matrix. Hence, computing the inverse can be avoided easily by applying the matrix inversion lemma. When the network has reached consensus, it holds that $\boldsymbol{z}_r = \boldsymbol{w}_r, \forall r \in \mathcal{R}$. The parameter $\varepsilon > 0$ can be tuned to control the distributed regression to prioritize consensus over minimizing the data fit in (13a) and vice versa. Furthermore, $\boldsymbol{v}_{r\ell}$ is a Lagrange multiplier that accounts for the consensus constraint (13b) and enables convergence to a consensus solution over the network. In terms of communication within the network, weight vectors $\boldsymbol{w}_r^{[n]}$, Lagrange multipliers $\boldsymbol{v}_{r\ell}^{[n]}$ and auxiliary variables $\boldsymbol{z}_r^{[n+1]}$ need to be exchanged by each receiver r with its neighbors $\ell \in \mathcal{N}_r$ in each iteration n. With the weight vector $\boldsymbol{w}_r^{[n]}$ available locally each receiver r can approximate the residuals $\boldsymbol{\rho}$ via $\widehat{\boldsymbol{\rho}}_r^{[n]} = \boldsymbol{G}\boldsymbol{w}_r^{[n]}$.

Once estimates of the traveltime residuals $\boldsymbol{\rho}$ are available, we can perform a local tomography update per receiver r. To this end, each receiver r has its own individual subsurface model $m_r(\boldsymbol{x})$. Then, following (4) each receiver computes an individual adjoint-state variable $\lambda_{r,s}(\boldsymbol{x})$ using the corresponding estimated traveltime residual $\widehat{\boldsymbol{\rho}}_r$ and afterwards a TT gradient $m_{\Delta,r}(\boldsymbol{x})$ via (5). Finally, each receiver updates its local subsurface model $m_r(\boldsymbol{x})$. In terms of vector notation we have the model update at iteration k:

$$\boldsymbol{m}_r^{[k+1]} = \boldsymbol{m}_r^{[k]} - \alpha_{\mathrm{TT}}^{[k]} \boldsymbol{m}_{\Delta,r} \tag{15}$$

Using a distributed regression scheme in combination with the TT formulas (4) and (5) enables each receiver to image the subsurface locally. Note that the iteration index for D-TOMO is k and therefore different to the iteration index of the distributed regression which is n. The reason is that in each D-TOMO iteration k, the distributed regression (14) is performed iteratively over the iteration n. When the distributed regression has finished, the gradient computation and the model update are performed locally. A pseudo-code of D-TOMO can be found in [12].

C. Adapt-Then-Combine Full Waveform Inversion (ATC-FWI)

D-TOMO enables the reconstruction of a velocity model in a distributed fashion within the network of receivers. However, since D-TOMO relies on the eikonal equation and on traveltimes, the spatial resolution of the obtained images will be rather low since it is limited by the first Fresnel zone [29]. To enhance the spatial resolution of the subsurface images FWI can be used. To this end, we proposed a distributed FWI based on the adapt-then-combine (ATC) technique [30] in [13]. In the following, we briefly review the algorithm here.

To enable a distributed implementation of the FWI, we first separate the original cost (7) over the receivers:

$$\mathcal{J}_\mathrm{FWI}(m) = \sum_{r \in \mathcal{R}} \mathcal{J}_{\mathrm{FWI},r}(m) \tag{16}$$

where

$$\mathcal{J}_{\mathrm{FWI},r}(m) = \frac{1}{2} \sum_{s=1}^{N_\mathrm{S}} \int_0^\tau \left(d_{s,r}^{\mathrm{syn}}(t,m) - d_{s,r}^{\mathrm{obs}}(t) \right)^2 \mathrm{d}t \tag{17}$$

is the local FWI cost at receiver r. For each local cost $\mathcal{J}_{\mathrm{FWI},r}$ a local gradient $m_{\Delta,r}(\boldsymbol{x})$ can be computed w.r.t. m following the same procedure as in the centralized FWI in Section II-B. To this end, forward wavefield $u_{s,r}(\boldsymbol{x},t)$ and adjoint wavefield $q_{s,r}(\boldsymbol{x},t)$ are required by each receiver r. The forward wavefield $u_{s,r}(\boldsymbol{x},t)$ is generated using the current local subsurface model $m_r(\boldsymbol{x})$ of receiver r whereas the adjoint wavefield uses the receiver-specific data residual only, i.e., $d_{s,r}^{\mathrm{syn}}(t,m) - d_{s,r}^{\mathrm{obs}}(t)$. The adjoint wavefield can be computed analogously following (11) with the local data residual placed at receiver position $\boldsymbol{x}_r$. Using the adjoint wavefield $q_{s,r}(\boldsymbol{x},t)$, each receiver r then computes the gradient following (10) where $m(\boldsymbol{x})$ is replaced by the current local subsurface model $m_r(\boldsymbol{x})$. To obtain subsurface images at each receiver as estimates of the global FWI result, the ATC technique is applied. To this end, velocity model and gradient are first discretized and then described by vectors $\boldsymbol{m}_r^{[k]}$ and $\boldsymbol{m}_{\Delta,r}^{[k]}$, respectively. Applying the ATC results in the following update formulas for receiver r and iteration k:

$$\text{(Adapt)} \quad \widetilde{\boldsymbol{m}}_r^{[k+1]} = \boldsymbol{m}_r^{[k]} - \alpha_{\mathrm{FWI}}^{[k]} \sum_{\ell \in \mathcal{N}_r} c_{\ell r} \boldsymbol{m}_{\Delta,\ell}^{[k]} \tag{18a}$$

$$\text{(Combine)} \quad \boldsymbol{m}_r^{[k+1]} = \sum_{\ell \in \mathcal{N}_r} a_{\ell r} \widetilde{\boldsymbol{m}}_\ell^{[k+1]} \tag{18b}$$

In each iteration k, first the gradients $\boldsymbol{m}_{\Delta,\ell}^{[k]}$ are exchanged among neighboring receivers $\ell \in \mathcal{N}_r$ and then fused to adapt the current local model $\boldsymbol{m}_r^{[k]}$ with a gradient descent step. After that, the intermediate models $\widetilde{\boldsymbol{m}}_r^{[k+1]}$ are exchanged among the receivers and then fused to give the updated model $\boldsymbol{m}_r^{[k+1]}$. The coefficients $c_{\ell r}, a_{\ell r} \in \mathbb{R}$ are weighting factors for the fusion process at receiver r w.r.t. data from neighbor ℓ. In [31] several selection strategies are proposed for the weights to enable convergence close to the global model. Equations (18) form the so-called ATC-FWI. A pseudo-code of the algorithm and further details can be found in [13].

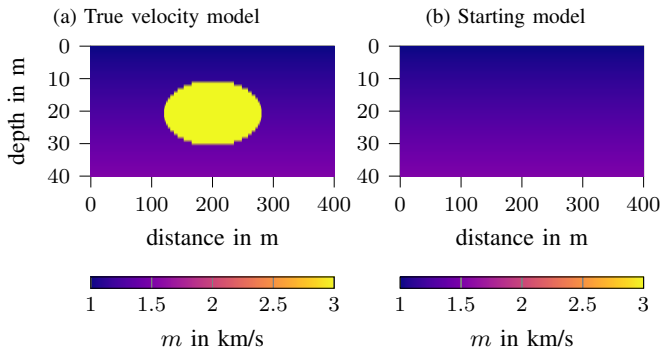


Fig. 4: (a) True velocity model of the elliptic anomaly and (b) background velocity as starting model.

D. Joint Imaging Scheme

As described in the previous sections both D-TOMO and ATC-FWI enable distributed imaging of the subsurface. However, D-TOMO alone obtains low-resolution subsurface images only due to the limited information contained in first-arrival traveltimes. On the other hand, ATC-FWI requires a sufficiently good initial model of the subsurface in order to converge to a reasonable solution. This is due to the fact that the cost function in FWI has multiple local minima in which the imaging scheme can get trapped. Hence, to enable high-resolution distributed subsurface imaging starting from a simple starting model we propose to use D-TOMO and ATC-FWI in a cascaded manner as illustrated in Fig. 2. It should be noted that images provided by D-TOMO can slightly differ among the receivers depending on their position and the connectivity of the seismic network. Hence, for the ATC-FWI each receiver is possibly initialized with a slightly different starting model. However, in the numerical evaluations we show that subsurface images provided by D-TOMO serve as sufficiently good starting models for the ATC-FWI to obtain high-resolution subsurface images. Additionally, as we show in Section IV-C in a joint scheme it is possible to perform D-TOMO for a low number of iterations $(1-5)$ only to obtain rough starting models. ATC-FWI is then still able to obtain sufficiently good subsurface images.

Remark: In both D-TOMO and ATC-FWI the same step size is used for each receiver in the model updates (15) *and* (18a)*, respectively. In general, a line search method can be used to optimize the step size value. However, this is only possible in a centralized setting where the global cost $\mathcal{J}_{\mathrm{TT}}$ and $\mathcal{J}_{\mathrm{FWI}}$ can be evaluated for different step size values. In a distributed setting, the global cost is not available locally at each receiver. Therefore, in D-TOMO and ATC-FWI the global cost needs to be estimated by each receiver to apply line search methods. Since this is out of the scope of this paper, we rely on a basic gradient descent update where the step size at each receiver is the same and follows a fixed strategy.*

IV. Numerical Evaluation

In this section, we present extensive numerical results of our proposed joint imaging scheme. We start with two synthetic velocity models before showing results using real seismic data acquired from measurements over a highway tunnel.

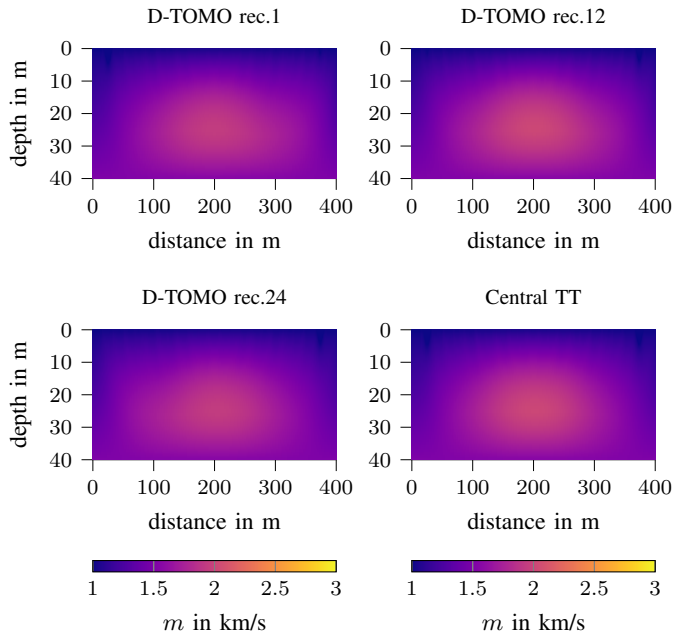


Fig. 5: D-TOMO imaging results for elliptic anomaly.

TABLE I: Parameter values for imaging the ellipse.

D-TOMO	$N_{\mathrm{TOMO}} = 10, \alpha^{[0]}_{\mathrm{TOMO}} = 250,$ $N_{\mathrm{ADMM}} = 100, \nu = 10^6, \varepsilon = 100, \sigma = 1$
ATC-FWI	$N_{\mathrm{FWI}} = 20, \alpha^{[0]}_{\mathrm{FWI}} = 0.5, \tau = 1\,\mathrm{s}, f_{\mathrm{d}} = 20\,\mathrm{Hz}$

A. Synthetic Elliptic Model

We consider a synthetic model consisting of an elliptic anomaly. At the first stage, D-TOMO obtains a smooth P-wave velocity model which is then refined by ATC-FWI. The true velocity model is depicted in Fig. 4a with the elliptic anomaly in the center and an increasing velocity as background model. As starting model for D-TOMO we assume knowledge of the background velocity, cf. Fig. 4b. We use $N_{\mathrm{R}} = 24$ receivers uniformly spread over the surface in a line array. Each receiver has a maximum of two neighbors to their left- and right-hand side. We employ $N_{\mathrm{S}} = 16$ shot positions which are again uniformly spread over the surface. As source signal we use a Ricker wavelet with dominant frequency $f_{\mathrm{d}} = 20\,\mathrm{Hz}$. The measurement time is set to $\tau = 1\,\mathrm{s}$.

The employed parameters for D-TOMO and ATC-FWI such as step size $\alpha_{\mathrm{TOMO}}, \alpha_{\mathrm{FWI}}$, no. of iterations $N_{\mathrm{TOMO}}, N_{\mathrm{ADMM}}, N_{\mathrm{FWI}}$, regularization parameters ν, ε and kernel bandwidth σ are listed in Table I. To improve convergence behavior to a local minimum, in both methods the step size is decaying exponentially over the iterations k via $\alpha^{[k]}_{\mathrm{TT}} = \alpha^{[0]}_{\mathrm{TT}} 0.95^k$ and $\alpha^{[k]}_{\mathrm{FWI}} = \alpha^{[0]}_{\mathrm{FWI}} 0.8^k$, respectively. In the model update we normalize the gradient to be in the range $[-1, 1]$ which is valid in our implementation due to a decaying step size. Fig. 5 shows the imaging results obtained

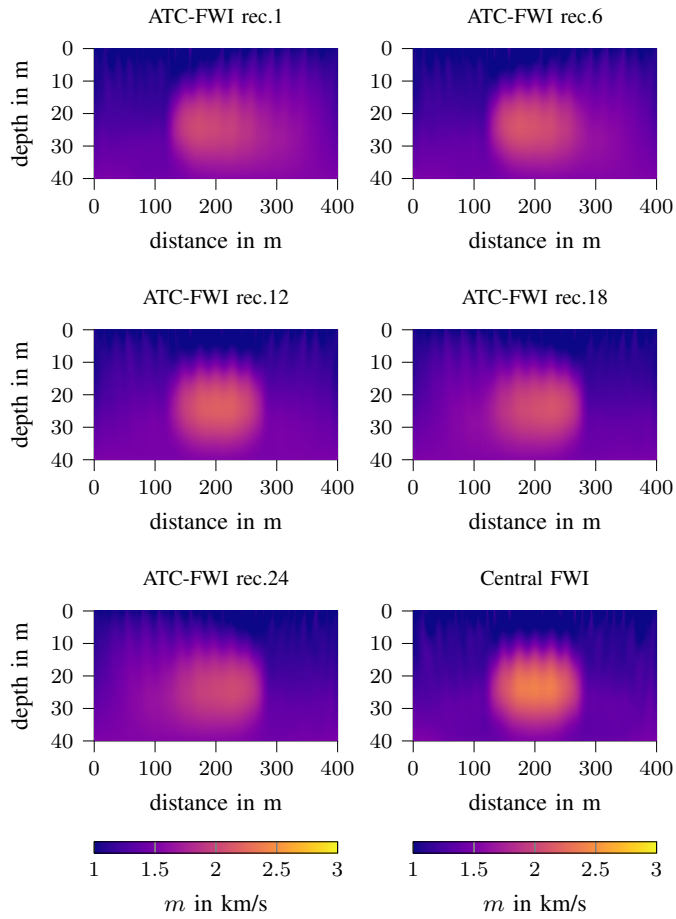


Fig. 6: ATC-FWI imaging results for elliptic anomaly.

by D-TOMO at three different receivers in the line array. It can be seen that D-TOMO is able to image the anomaly in the subsurface with similar quality over the shown receivers. However, the spatial resolution is low. Thus the contour of the anomaly cannot be identified accurately. To enhance the spatial resolution we now apply the ATC-FWI. We use the subsurface models obtained at each receiver by D-TOMO as starting models at the corresponding receivers for ATC-FWI. The resulting subsurface images for five receivers are depicted in Fig. 6. It can be clearly seen that the initial model obtained by D-TOMO is significantly improved in its spatial resolution by ATC-FWI. Compared to the true velocity model we observe high accuracy at all receivers especially at the middle receiver no. 12. However the absolute velocity of $3\,\mathrm{km/s}$ in the anomaly is not recovered exactly. Despite that the reconstructed models show high similarity over the receivers and achieve similar reconstruction performance compared to central FWI. For receivers 1 and 24 (first and last in the array) one observes slight deviations in the right- and left-hand side of the domain, respectively. This is due to the location of the respective receiver. For receiver 1, being the first on the left-hand side of the line array, deviations can be observed on the right-hand side since seismic data from receivers here will likely diminish in the data exchange process

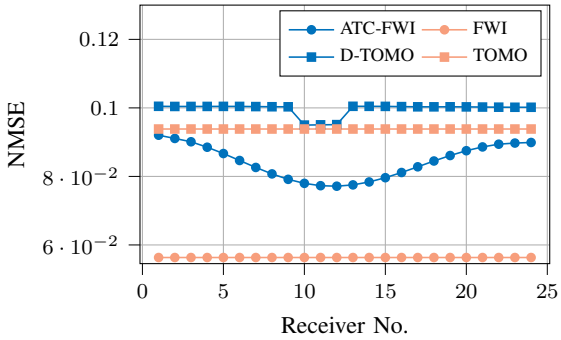


Fig. 7: NMSE over receiver no. for D-TOMO and ATC-FWI for elliptic anomaly after $N_{\mathrm{TOMO}} = 10$ and $N_{\mathrm{FWI}} = 20$ iterations, respectively.

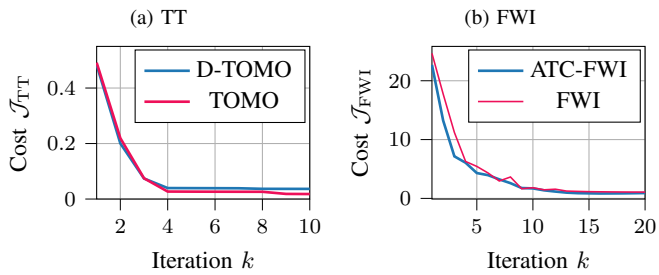


Fig. 8: Costs of (a) D-TOMO vs. TOMO and (b) ATC-FWI vs. FWI for elliptic model.

through the network. The same behavior can be observed for the last receiver on the far right-hand side of the array. For receivers 6 and 18 the deviations are smaller.

In Fig. 7 the $\mathrm{NMSE} = ||\boldsymbol{m}_r - \boldsymbol{m}_{\mathrm{true}}||_2^2 / ||\boldsymbol{m}_{\mathrm{true}}||_2^2$ over the receiver no. is depicted for both D-TOMO and ATC-FWI. As reference, the NMSE performance of the centralized schemes is shown as well. One can observe that especially receivers located in the center of the array achieve lowest NMSE. On one hand, this is due to the higher connectivity of these receivers in the network. On the other hand, their position right above the center of the anomaly is beneficial for their reconstruction performance. Fig. 8a and 8b show the residual cost over the iterations for both D-TOMO and ATC-FWI compared to their central counterpart. In both cases, we observe that the costs converge over the iterations and that our proposed distributed imaging methods achieve costs close to their central counterpart.

B. Synthetic SEG/EAGE Salt Model

As a further synthetic example we consider the SEG/EAGE salt model [16] that is frequently used in seismic imaging as a benchmark model. For our purposes we adapt the original 3D salt model by selecting a 2D slice, scaling the velocities

TABLE II: Parameters for imaging the SEG/EAGE salt model.

D-TOMO	$N_{\mathrm{TOMO}} = 10, \alpha_{\mathrm{TOMO}}^{[0]} = 150,$ $N_{\mathrm{ADMM}} = 100, \nu = 10^4, \varepsilon = 100, \sigma = 1$
ATC-FWI	$N_{\mathrm{FWI}} = 15, \alpha_{\mathrm{FWI}}^{[0]} = 0.3, \tau = 1\,\mathrm{s}, f_{\mathrm{d}} = 20\,\mathrm{Hz}$

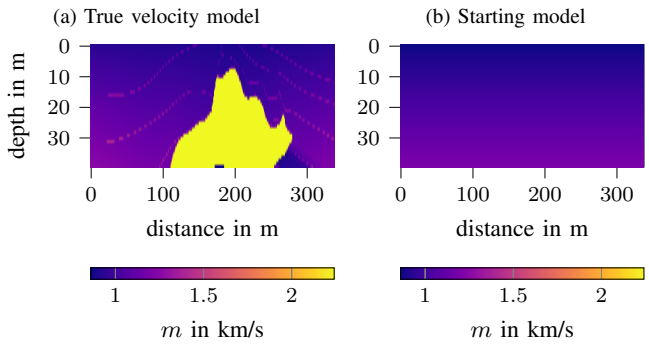


Fig. 9: (a) True velocity model of the salt model and (b) velocity gradient as starting model.

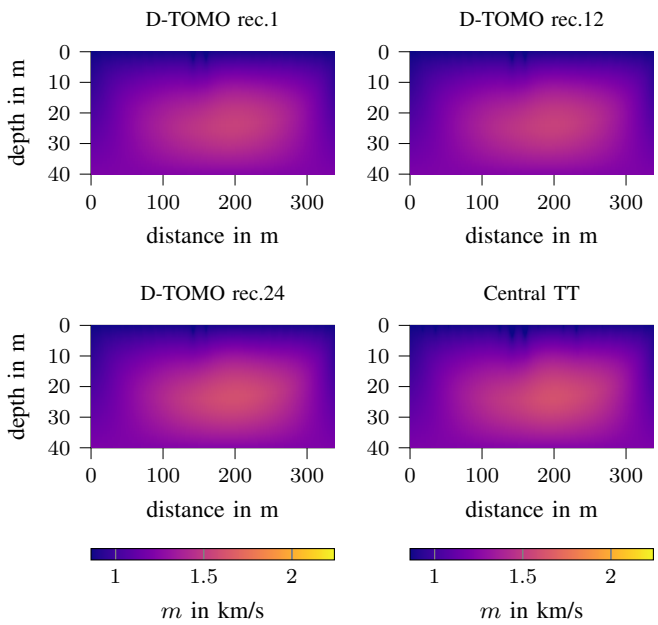


Fig. 10: D-TOMO results for salt model.

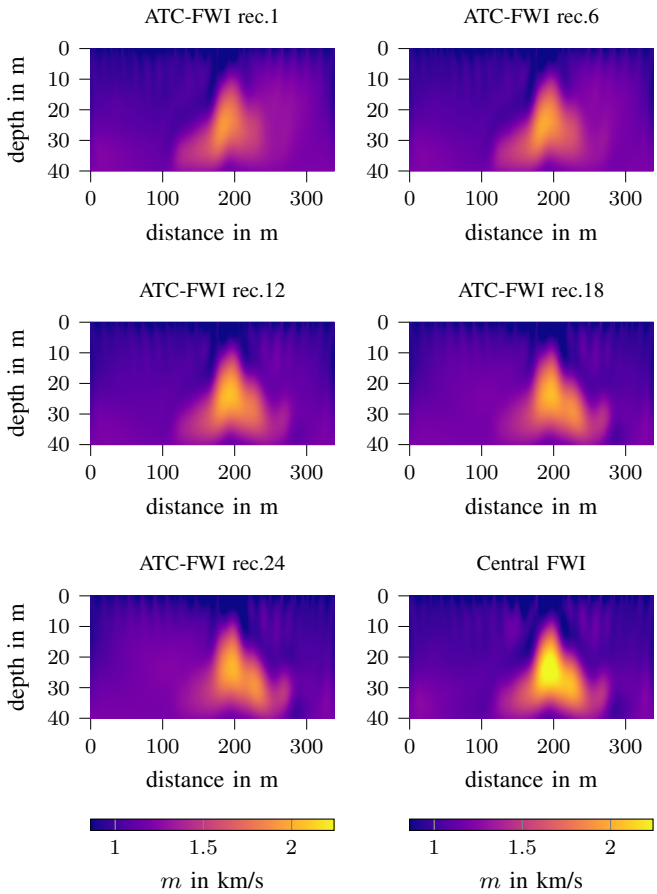


Fig. 11: Imaging results for salt model using ATC-FWI and FWI.

and resampling the grid points to obtain a smaller model. The resulting ground truth model for our experiment is depicted in Fig. 9a. Again we apply D-TOMO first and then ATC-FWI. For the starting model of D-TOMO we assume an increasing velocity model as shown in Fig. 9b that resembles the background velocity of the true model. We use the same array configuration as in Section IV-A except that each receiver has a maximum of three receivers to their left- and right-hand side, respectively. We employ $N_{\mathrm{S}} = 20$ shot positions uniformly aligned over the surface.

For D-TOMO and ATC-FWI the respective parameters are listed in Table II. As before, the step size is decaying exponentially in the same fashion as described in Section IV-A. For D-TOMO the resulting subsurface images are depicted in Fig. 10. One can observe that the anomaly in the subsurface is recovered by D-TOMO, however, with very low spatial resolution. The salt anomaly from the true model is barely visible. Nevertheless, the individual subsurface images of the receivers coincide very well with the central result. Fig. 11 shows the images recovered by ATC-FWI when using the results from D-TOMO as respective starting models. It can be clearly seen that the low-resolved anomaly from D-TOMO is significantly enhanced in its spatial resolution for the depicted receivers. In particular, the salt anomaly from the true velocity model can now be identified more clearly especially for receivers 12, 18 and 24. Compared to the central FWI result, ATC-FWI achieves similar imaging results for receiver 12, 18 and 24. At receiver 1 and 6, deviations to the central result are more visible since data from receivers on the right-hand side of the line array diminish over the data exchange process in the network. However, these data are decisive to reconstruct the finer structures on the right-hand side of the salt anomaly. Therefore, details of the anomaly from $x = 200\,\mathrm{m}$ on are not recovered accurately. This is also visible from the NMSE in Fig. 12: Receivers located in the center or right-hand side of the array achieve a lower NMSE compared to receivers on the left-hand side for both D-TOMO and ATC-FWI. Fig. 13a and 13b depict the costs over the iterations for both D-TOMO and ATC-FWI and their central benchmarks. The curves confirm that both distributed schemes perform very close to their central benchmarks. Also for the SEG/EAGE salt model our proposed joint imaging scheme can be used to image such an anomaly in a distributed fashion.

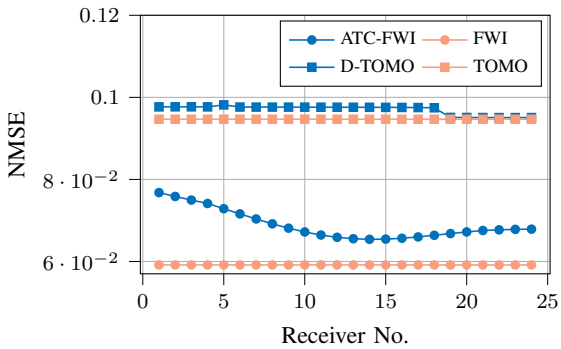


Fig. 12: NMSE over receiver no. for D-TOMO and ATC-FWI for salt model after $N_{\mathrm{TOMO}} = 20$ and $N_{\mathrm{FWI}} = 20$ iterations.

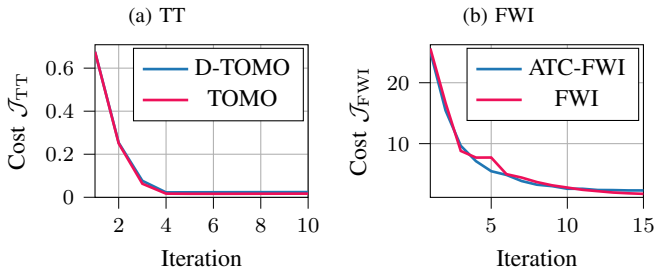


Fig. 13: Costs of (a) D-TOMO vs. TOMO and (b) ATC-FWI vs. FWI for SEG/EAGE salt model.

C. Joint Imaging Performance With Few Iterations of D-TOMO

In the following, we investigate the effect of low iterations for D-TOMO on the imaging performance of ATC-FWI. Since D-TOMO consists of two nested iterative methods (the ADMM for the distributed regression and the tomographic updates themselves) it is desirable to keep the number of iterations low to reduce communication exchanges among the receivers. To this end, we perform D-TOMO for $N_{\mathrm{TOMO}} = 1, \ldots, 5$ iterations. The obtained images are then used to initialize ATC-FWI which is run for $N_{\mathrm{FWI}} = 20$ iterations. We perform the evaluation for both ellipse and salt model with the same set of parameters used in the respective sections.

Fig. 14 depicts the imaging results over the number of iterations performed by D-TOMO. To keep the amount of images limited we show the results of receiver 18. The numbers in the boxes indicate the NMSE of the recovered images. It is noticeable that even with only one iteration performed by D-TOMO, ATC-FWI is still able to obtain good imaging results. With an increased number of iterations the quality is enhanced. Compared to $N_{\mathrm{TOMO}} = 10$ iterations as performed in the preceding evaluations, ATC-FWI achieves satisfactory imaging results. For the salt model, with higher number of iterations for D-TOMO the imaging results by ATC-FWI contain higher velocity values for the salt body that are closer to the ground truth, cf. Fig. 11. Nevertheless, these results demonstrate that with only a low number of iterations for D-TOMO, ATC-FWI is robust enough to still obtain high-resolution subsurface images with satisfactory quality.

TABLE III: Parameters for imaging the highway tunnel.

D-TOMO	$N_{\mathrm{TOMO}} = 20, \alpha^{[0]}_{\mathrm{TOMO}} = 250,$ $N_{\mathrm{ADMM}} = 100, \nu = 10^4, \varepsilon = 100, \sigma = 1$
ATC-FWI	$N_{\mathrm{FWI}} = 20, \alpha^{[0]}_{\mathrm{FWI}} = 0.01, \tau = 0.4\,\mathrm{s}, f_{\mathrm{d}} = 20\,\mathrm{Hz}$

D. Field Data Over Highway Tunnel

Lastly, we apply our proposed joint distributed imaging scheme to real seismic data acquired in the field. Specifically, we use seismic data that we recorded over a tunnel on the A99 highway in Aubing, Germany. Here, we used $N_{\mathrm{R}} = 16$ GS-20DX geophones from Geospace Technologies with a resonance frequency of $40\,\mathrm{Hz}$. These were placed in a line array with a distance of $3\,\mathrm{m}$ to each other such that the complete array had a length of $45\,\mathrm{m}$. The array has been distributed over the highway tunnel where a part of the array covers the tunnel, see Fig. 16. This is done to identify the tunnel structure more easily in the reconstructed images. Furthermore, $N_{\mathrm{S}} = 8$ shot positions were used with four shots per position to stack the acquired seismic data for a higher signal-to-noise ratio in the measurements. As source signal we used strikes with a $4\,\mathrm{kg}$ sledgehammer. In total, four measurement phases were conducted. In each phase the complete line array was moved such that it had a $3\,\mathrm{m}$ distance to its preceding location and new measurements were performed. Fig. 15 exemplary shows measured seismic traces at four different shot positions. One can see that the shot is moved into positive x-direction and multiple refracted waves are recorded. The data was recorded using a Geode Seismograph from Geometrics with 16 channels. The measurement data has been published as an open data set (see also [32]).

To obtain first-arrival traveltimes from our measurements we pick them manually. Then we apply D-TOMO on the picked traveltime data. For the depicted results we use data from the first array (Line 1 in Fig. 16). We assume that each receiver has two neighbors to their left- and right-hand side except for the first and last receiver who have one neighbor only. The resulting images at three receivers and the central result are shown in Fig. 17. Since the spatial resolution of D-TOMO is low an accurate identification of the tunnel edges is difficult. It should be noted that our measurement data is likely to contain significant noise due to cars driving through the tunnel during the measurements. Nevertheless, the outer tunnel shape is clearly visible in the reconstructed images. From the highway management site we knew that the tunnel ceiling is expected to lie at a depth of approximately $4-6\,\mathrm{m}$. Moreover, GPS data indicates the start of the tunnel from the first geophone in the line array to be at approximately $18\,\mathrm{m}$. Also here the subsurface images obtained by D-TOMO coincide well with these data. Compared to the central imaging result, D-TOMO obtains very similar results.

Now we apply ATC-FWI to the tunnel data to increase the resolution of the images. To this end, we assume a Ricker wavelet as source with a dominant frequency of $f_{\mathrm{d}} = 20\,\mathrm{Hz}$. The parameters used are summarized in Table III and the resulting images are depicted in Fig. 18. Compared to D-TOMO, we observe some enhancement in the spatial reso-

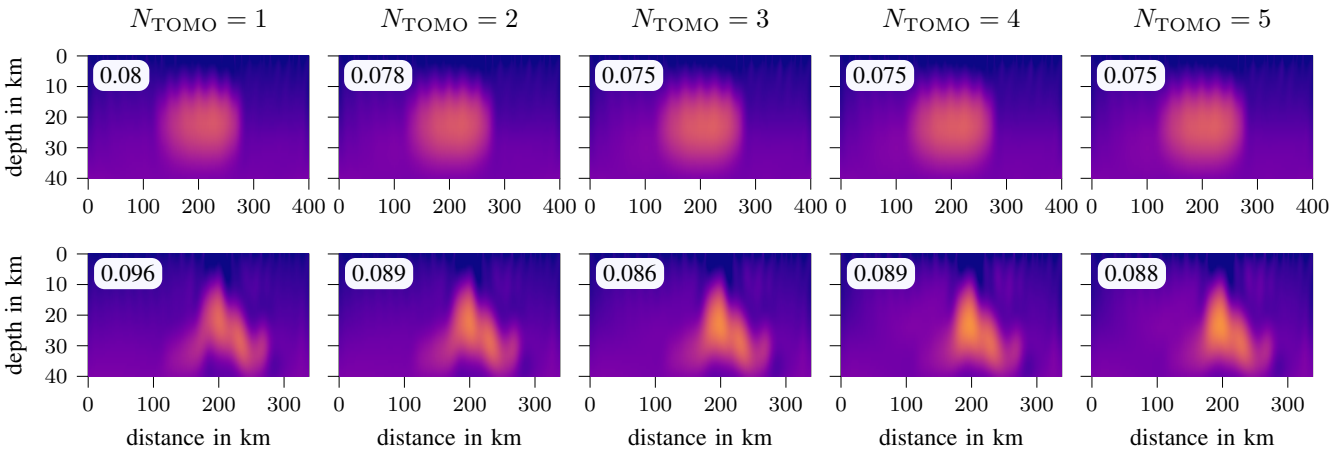
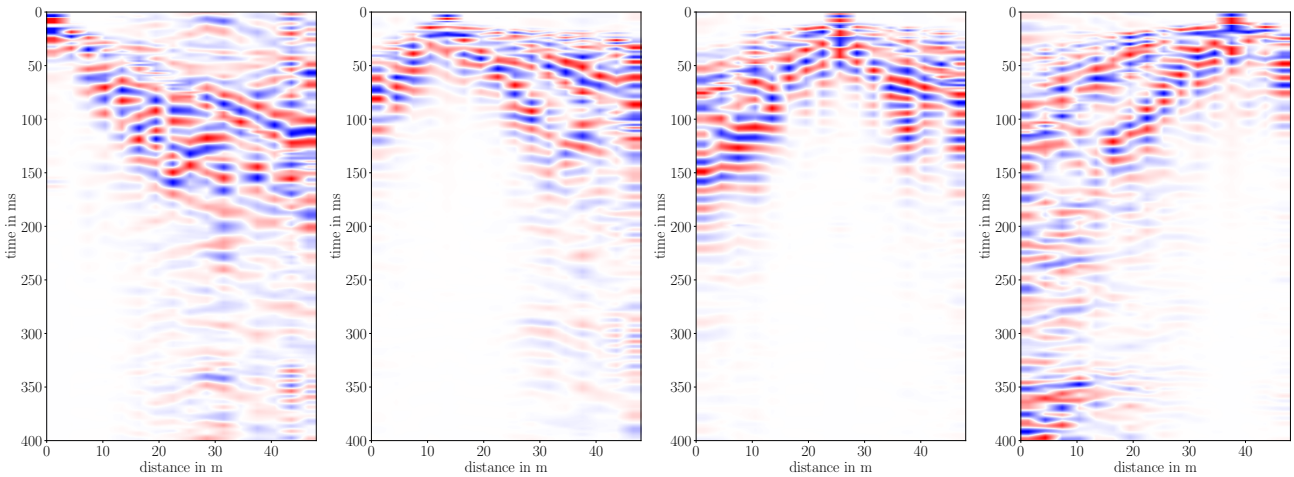


Fig. 14: Imaging results obtained by ATC-FWI after $N_{\mathrm{FWI}} = 20$ iterations for a varying number of iterations N_{TOMO} employed by D-TOMO. Depicted are the images at rec. 18 for (row 1) the ellipse and (row 2) the salt model. Numbers in the boxes are the NMSE values.



(a) Shot no. 1 (b) Shot no. 3 (c) Shot no. 5 (d) Shot no. 7

Fig. 15: Seismic traces of measurement data over highway tunnel for four selected shot positions. The data is normalized between $[-1, 1]$.

lution. However, a significant improvement as in the synthetic examples cannot be observed. A possible reason for this behavior are low frequency parts that are missing in the recorded signals. Low frequencies in seismic measurements are decisive for FWI to achieve a high resolution. Fig. 19 depicts the power spectral density (PSD) of the recorded signals averaged over all receivers for one shot. It can be seen that most of the power is located roughly around 50 Hz. This is due to a high resonance frequency of 40 Hz in the geophones and also due to using hammer strikes that possibly do not excite low frequencies in the area of 10 Hz. Another reason can lie in the simplified wave equation that we employ here. Since we use the acoustic wave equation we are able to model P-wave propagation only ignoring S-wave propagation. Hence, possible shear wave components in the recorded signals are not exploited by our algorithms such that the image by D-TOMO is not significantly enhanced by ATC-FWI. Nevertheless, our results demonstrate that D-TOMO can be successfully applied to perform imaging on real seismic data. For ATC-FWI to be successful low frequencies are required in the seismic data.

V. Conclusion

In this paper, we proposed and investigated a joint imaging scheme consisting of D-TOMO and ATC-FWI for distributed subsurface imaging in seismic networks. By means of distributed imaging each receiver in the network is able to obtain an image of the subsurface locally by exploiting information exchange with neighboring receivers. Our proposed joint imaging scheme is fully distributed, i.e., it requires neither a full mesh topology nor anchor nodes that process intermediate results. Thus, our scheme has high flexibility in terms of topology and each receiver can be designed in the same fashion. In extensive numerical evaluations we illustrated that our proposed joint scheme achieves central imaging performance. Images obtained by D-TOMO can be directly used to initialize ATC-FWI. We observed that ATC-FWI is able to enhance

Fig. 16: Google maps view of tunnel over A99 highway in Aubing, Germany, and geophone positions. The tunnel is indicated in yellow.

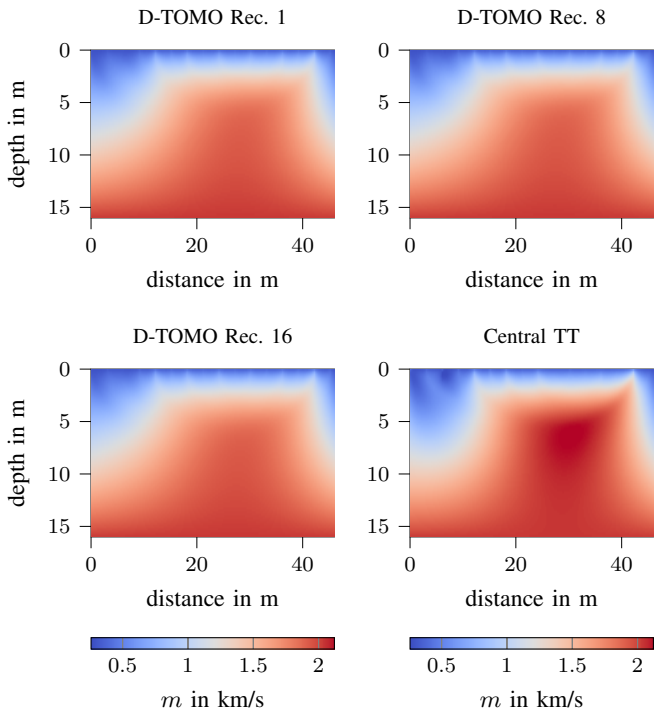


Fig. 17: D-TOMO and TT results for highway tunnel data after $N_{\text{TOMO}} = 20$ iterations.

the imaging results by D-TOMO significantly in their spatial resolution especially for the synthetic models. In particular, running D-TOMO for only a low number of iterations showed to be sufficient for ATC-FWI to still obtain high-resolution subsurface images. For real seismic data we used an own dataset that was recorded over a highway tunnel. Here, D-TOMO is able to recover rough structures of the tunnel that coincide with available ground truth information. However, ATC-FWI does not significantly enhance the imaging results in the used setting. Possible reasons are the high resonance frequency of the used geophones and usage of the simplified acoustic wave equation that considers P-wave propagation only. In a future experiment geophones with lower resonance frequency of e.g. 10 Hz should be considered to obtain low

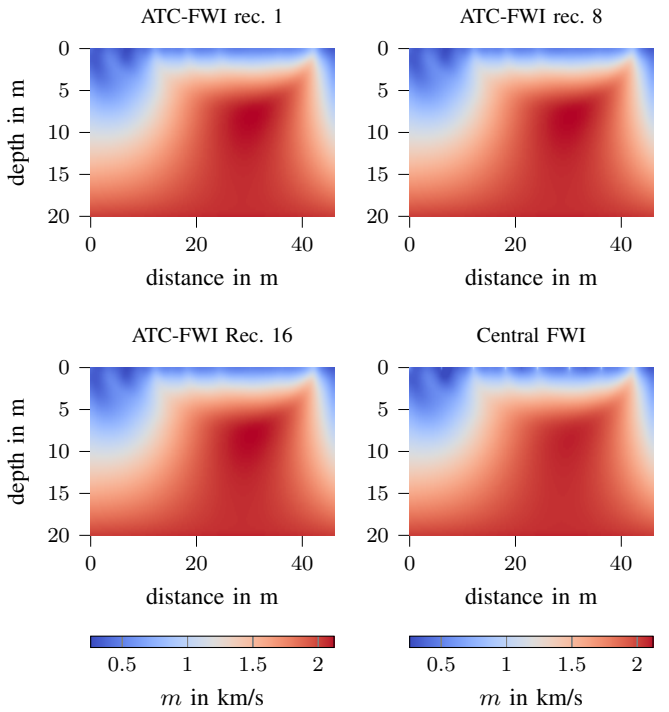


Fig. 18: ATC-FWI and FWI results for highway tunnel data after $N_{\text{FWI}} = 20$ iterations.

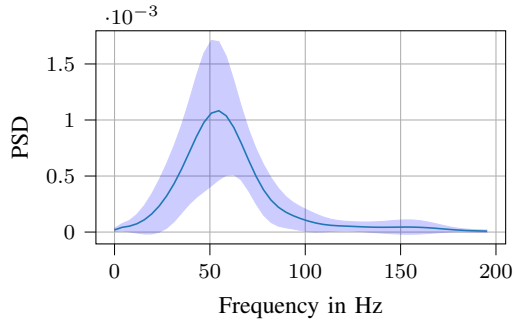


Fig. 19: Power spectral density of the recorded seismic signals averaged over the receivers. Most power is located around 50 Hz indicating that low frequencies are missing in the recorded traces.

frequency parts in the recorded signal that are decisive for FWI to obtain high-resolution images.

Appendix

A. Derivation of adjoint-state for traveltime tomography

To apply the adjoint-state method to traveltime tomography, the original cost function in (1) is extended as follows [23]:

$$\mathcal{L}(m, \lambda_s, T_s) = \frac{1}{2} \sum_{s=1}^{N_S} \int_{\partial\Omega} |T_s(m, \boldsymbol{x}) - T_s^{\text{obs}}(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x} - \frac{1}{2} \sum_{s=1}^{N_S} \int_{\Omega} \lambda_s(\boldsymbol{x}) \left(|\nabla T_s(\boldsymbol{x})|^2 - \frac{1}{m(\boldsymbol{x})^2} \right) \mathrm{d}\boldsymbol{x}, \tag{19}$$

where $\lambda_s(\boldsymbol{x})$ is the adjoint-state variable. It can be interpreted as a Lagrangian variable that explicitly accounts for

the constraint given by the eikonal equation (2). Following Lagrangian optimization the partial derivatives of $\mathcal{L}(m, \lambda_s, T_s)$ w.r.t. λ_s and T_s need to vanish. However, both λ_s and T_s are functions w.r.t. $\boldsymbol{x}$. Therefore, we need to make use of calculus of variations and compute the functional derivative of $\mathcal{L}$ accordingly. To this end, we cast (19) into the form

$$\mathcal{L}(T_s) = \int_{\partial\Omega} g(T_s, \boldsymbol{x})\mathrm{d}\boldsymbol{x} - \int_{\Omega} f(T_s, \nabla T_s, \boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{20}$$

with

$$g(T_s, \boldsymbol{x}) = \frac{1}{2}\sum_{s=1}^{N_\mathrm{S}} \left|T_s(m, \boldsymbol{x}) - T_s^{\mathrm{obs}}(\boldsymbol{x})\right|^2 \tag{21a}$$

$$f(T_s, \nabla T_s, \boldsymbol{x}) = \frac{1}{2}\sum_{s=1}^{N_\mathrm{S}} \lambda_s(\boldsymbol{x}) \left(|\nabla T_s(\boldsymbol{x})|^2 - \frac{1}{m(\boldsymbol{x})^2}\right). \tag{21b}$$

For the time being, we consider the synthesized travel time T_s for shot s as the variable of interest. For the other functions m and λ_s derivation of the functional derivative follows analogously. Now, assume a perturbation δ of (20) [33]:

$$\begin{aligned}\delta\mathcal{L}(T_s) &= \int_{\partial\Omega} \delta g(T_s)\,\mathrm{d}\boldsymbol{x} - \int_{\Omega} \delta f(T_s, \nabla T_s)\,\mathrm{d}\boldsymbol{x} \\ &= \int_{\partial\Omega} \left[\frac{\partial g}{\partial T_s}\delta T_s\right]\mathrm{d}\boldsymbol{x} - \\ &\quad \int_{\Omega} \left[\frac{\partial f}{\partial T_s}\delta T_s + \frac{\partial f}{\partial(\nabla T_s)}\cdot\nabla\delta T_s\right]\mathrm{d}\boldsymbol{x}\end{aligned} \tag{22}$$

For the last term in the equation above we apply the product rule for the divergence:

$$\frac{\partial f}{\partial(\nabla T_s)}\cdot\nabla\delta T_s = \nabla\cdot\left(\frac{\partial f}{\partial(\nabla T_s)}\delta T_s\right) - \delta T_s\nabla\cdot\left(\frac{\partial f}{\partial(\nabla T_s)}\right)$$

Then we arrive at

$$\begin{aligned}\delta\mathcal{L}(T_s) = \int_{\partial\Omega} \frac{\partial g}{\partial T_s}\delta T_s\,\mathrm{d}\boldsymbol{x} - \int_{\Omega}\Bigg[\frac{\partial f}{\partial T_s}\delta T_s \\ + \nabla\cdot\left(\frac{\partial f}{\partial(\nabla T_s)}\delta T_s\right) - \nabla\cdot\left(\frac{\partial f}{\partial(\nabla T_s)}\right)\Bigg]\mathrm{d}\boldsymbol{x}.\end{aligned} \tag{23}$$

Now we apply the divergence theorem as follows:

$$\int_{\Omega}\nabla\cdot\left(\frac{\partial f}{\partial(\nabla T_s)}\delta T_s\right)\mathrm{d}\boldsymbol{x} = \int_{\partial\Omega}\left(\frac{\partial f}{\partial(\nabla T_s)}\cdot\boldsymbol{n}\right)\delta T_s\mathrm{d}\boldsymbol{x}, \tag{24}$$

where $\boldsymbol{n}$ is the unit vector normal to the boundary $\partial\Omega$. Then we finally arrive at

$$\begin{aligned}\delta\mathcal{L}(T_s) = \int_{\partial\Omega}\left[\frac{\partial g}{\partial T_s} - \frac{\partial f}{\partial(\nabla T_s)}\cdot\boldsymbol{n}\right]\delta T_s\mathrm{d}\boldsymbol{x} \\ - \int_{\Omega}\left[\frac{\partial f}{\partial T_s} - \nabla\cdot\left(\frac{\partial f}{\partial(\nabla T_s)}\right)\right]\delta T_s\mathrm{d}\boldsymbol{x}\end{aligned} \tag{25}$$

We now require $\partial\mathcal{L}/\partial T_s = 0$. With the *fundamental lemma of calculus of variations* we can set the integrands to zero [33]:

$$\nabla\cdot\left(\frac{\partial f}{\partial(\nabla T_s)}\right) - \frac{\partial f}{\partial T_s} = 0 \qquad \text{in } \Omega \tag{26a}$$

$$\frac{\partial g}{\partial T_s} - \frac{\partial f}{\partial(\nabla T_s)}\cdot\boldsymbol{n} = 0 \qquad \text{on } \partial\Omega \tag{26b}$$

Now we can evaluate the partial derivatives by inserting $g(T_s)$ and $f(T_s, \nabla T_s)$ from (21) and treating both T_s and ∇T_s as variables. This gives the following set of equations:

$$\nabla\cdot(\lambda_s(\boldsymbol{x})\nabla T_s(\boldsymbol{x})) = 0, \qquad \boldsymbol{x}\in\Omega \tag{27a}$$

$$\lambda_s(\boldsymbol{x})\left(\nabla T_s(\boldsymbol{x})\cdot\boldsymbol{n}\right) = T_s(\boldsymbol{x}) - T_s^{\mathrm{obs}}(\boldsymbol{x}), \quad \boldsymbol{x}\in\partial\Omega \tag{27b}$$

By assuming that receivers are placed on the boundary $\partial\Omega$ we obtain the adjoint-state equations as given in (4).

B. Derivation of gradient for traveltime tomography

According to the adjoint-state method, to obtain the gradient or total derivative $\mathrm{d}\mathcal{J}(m)/\mathrm{d}m$ we need to compute the derivative $\partial\mathcal{L}(m)/\partial m$. For this partial derivative functional derivatives are used since m is a function over $\boldsymbol{x}$. Therefore, we apply the same procedure when calculating the adjoint-state equations but replace T_s by m in conditions (26) to obtain $\partial\mathcal{L}(m)/\partial m$.

Since the integrand over $\partial\Omega$ in $\mathcal{L}(m, \lambda_s, T_s)$ from (19) does not contain m explicitly, the partial derivative needs to be applied on the second summand only:

$$\begin{aligned}&\frac{\partial\mathcal{L}(m)}{\partial m} \\ &= \frac{\partial}{\partial m}\left\{-\frac{1}{2}\sum_{s=1}^{N_\mathrm{S}}\int_{\Omega}\lambda_s(\boldsymbol{x})\left(|\nabla T_s(\boldsymbol{x})|^2 - \frac{1}{m(\boldsymbol{x})^2}\right)\mathrm{d}\boldsymbol{x}\right\} \\ &= -\sum_{s=1}^{N_\mathrm{S}}\frac{\lambda_s(\boldsymbol{x})}{m(\boldsymbol{x})^3} = \frac{\mathrm{d}\mathcal{J}(m)}{\mathrm{d}m}\end{aligned} \tag{28}$$

which is the result for the gradient given in (3).

C. Derivation of gradient for full waveform inversion

We consider a least-squares cost between seismic measurements $d_{s,r}^{\mathrm{obs}}(t)$ and synthesized seismic data $d_{s,r}^{\mathrm{syn}}(t, m)$, cf. [15]:

$$\min_m \mathcal{J}_{\mathrm{FWI}}(m) = \frac{1}{2}\sum_{s=1}^{N_\mathrm{S}}\sum_{r=1}^{N_\mathrm{R}}\int_0^{\tau}\left(d_{s,r}^{\mathrm{syn}}(t, m) - d_{s,r}^{\mathrm{obs}}(t)\right)^2\mathrm{d}t \tag{29}$$

where $d_{s,r}^{\mathrm{syn}}(t, m) = S_{s,r}u_s(t, m)$. We apply the adjoint-state method and extend the cost function following Section 2.2. and 3.2 in [23]:

$$\begin{aligned}\mathcal{L}_{\mathrm{FWI}}(m) &= \frac{1}{2}\sum_{s,r}\int_0^{\tau}\left(d_{s,r}^{\mathrm{syn}}(t, m) - d_{s,r}^{\mathrm{obs}}(t)\right)^2\mathrm{d}t \\ &+ \sum_{s=1}^{N_\mathrm{S}}\int_0^{\tau}\Bigg\langle\lambda_s(\boldsymbol{x}, t), \frac{1}{m(\boldsymbol{x})^2}\frac{\partial^2 u_s(\boldsymbol{x}, t)}{\partial t^2} - \Delta u_s(\boldsymbol{x}, t) \\ &- f(\boldsymbol{x}, t)\Bigg\rangle_{\Omega}\mathrm{d}t + \sum_{s=1}^{N_\mathrm{S}}\left\langle\gamma_s^0, u_s(0)\right\rangle_{\Omega} + \sum_{s=1}^{N_\mathrm{S}}\left\langle\gamma_s^1, \frac{\partial u_s(0)}{\partial t}\right\rangle_{\Omega}\end{aligned} \tag{30}$$

with $\langle\lambda_s, u_s\rangle_{\Omega} = \int_{\Omega}\lambda_s(\boldsymbol{x})u_s(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ denoting the inner product between two functions in the domain space Ω and γ_s^0, γ_s^1 denoting the adjoint-state variables for the initial boundary conditions of $u_s(t)$. According to the adjoint-state method (cf.

Section 3.2 in [23]), we can compute the gradient of $\mathcal{J}_{\mathrm{FWI}}(m)$ w.r.t. m via

$$\frac{\mathrm{d}\mathcal{J}_{\mathrm{FWI}}(m)}{\mathrm{d}m} = \frac{\partial \mathcal{L}_{\mathrm{FWI}}(m)}{\partial m} \tag{31}$$

where $\mathrm{d}/\mathrm{d}m$ denotes the total derivative. Again, since m is a function, we use functional derivatives and the result from (26a). However, note that $\mathcal{L}_{\mathrm{FWI}}(m)$ does not depend on ∇m such that we can simply build the derivative of $\mathcal{L}_{\mathrm{FWI}}(m)$ by treating m as a variable. We then arrive at

$$\begin{aligned}\frac{\mathrm{d}\mathcal{J}_{\mathrm{FWI}}(m)}{\mathrm{d}m} &= \frac{\partial \mathcal{L}_{\mathrm{FWI}}(m)}{\partial m} \\ &= -\frac{2}{m(\boldsymbol{x})^3} \sum_{s=1}^{N_{\mathrm{S}}} \int_0^{\tau} \lambda_s(\boldsymbol{x},t) \frac{\partial^2 u_s(\boldsymbol{x},t)}{\partial t^2} \mathrm{d}t \end{aligned} \tag{32}$$

which is the result given in (10) with the substitution $\lambda_s(\boldsymbol{x},t) = q_s(\boldsymbol{x},\tau - t)$. A derivation of the adjoint-state equation to compute $\lambda_s(\boldsymbol{x},t)$ is given e.g. in [23].

REFERENCES

[1] D. Giardini, P. Lognonné, W. B. Banerdt *et al.*, "The seismicity of Mars," *Nature Geoscience*, vol. 13, no. 3, pp. 205–212, 2020.

[2] J. Lai, Y. Xu, R. Bugiolacchi *et al.*, "First look by the yutu-2 rover at the deep subsurface structure at the lunar farside," *Nature Communications*, vol. 11, no. 1, Jul. 2020.

[3] S. C. Stähler, A. Mittelholz, C. Perrin *et al.*, "Tectonics of cerberus fossae unveiled by marsquakes," *Nature Astronomy*, vol. 6, no. 12, pp. 1376–1386, Oct. 2022.

[4] P. F. Miaja, F. Navarro-Medina, D. G. Aller *et al.*, "Robocrane: A system for providing a power and a communication link between lunar surface and lunar caves for exploring robots," *Acta Astronautica*, vol. 192, pp. 30–46, 2022.

[5] S. W. Courville *et al.*, "ARES and Artemis: The Autonomous Roving Exploration System for Active Source Seismology on the Moon," in *Lunar Surface Science Workshop*, vol. 2241, May 2020.

[6] B.-S. Shin and D. Shutin, "Subsurface exploration on Mars and Moon with a robotic swarm," in *Global Space Exploration Conference*, 2021.

[7] G. Kamath, L. Shi, W.-Z. Song, and J. Lees, "Distributed travel-time seismic tomography in large-scale sensor networks," *Journal of Parallel and Distributed Computing*, vol. 89, pp. 50–64, 2016.

[8] F. Li, M. Valero, Y. Cheng, T. Bai, and W. Z. Song, "Distributed sensor networks based shallow subsurface imaging and infrastructure monitoring," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 241–250, 2020.

[9] S. Wang, F. Li, M. Panning, S. Tharimena, S. Vance, and W. Song, "Ambient noise tomography with common receiver clusters in distributed sensor networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 656–666, 2020.

[10] D. J. White, "Two-dimensional seismic refraction tomography," *Geophysical Journal International*, vol. 97, no. 2, pp. 223–245, 1989.

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

[12] B.-S. Shin and D. Shutin, "Distributed traveltime tomography using kernel-based regression in seismic networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[13] ——, "Adapt-then-combine full waveform inversion for distributed subsurface imaging in seismic networks," in *IEEE ICASSP*, 2021, pp. 4700–4704.

[14] Ò. C. Agudo, N. V. Da Silva, M. Warner, and J. Morgan, "Acoustic full-waveform inversion in an elastic world," *Geophysics*, vol. 83, no. 3, pp. R257–R271, 2018.

[15] W. Virieux, "An overview of full-waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, 2009.

[16] F. Aminzadeh, J. Brac, and T. Kunz, *3-D Salt and Overthrust models*. SEG/EAGE 3-D Modeling Series No.1, 1997.

[17] C. Taillandier, M. Noble, H. Chauris, and H. Calandra, "First-arrival traveltime tomography based on the adjoint-state method," *Geophysics*, vol. 74, no. 6, 2009.

[18] E. Treister and E. Haber, "A fast marching algorithm for the factored eikonal equation," *Journal of Computational Physics*, vol. 324, pp. 210–225, Nov. 2016.

[19] Y.-H. R. Tsai, L.-T. Cheng, S. Osher, and H.-K. Zhao, "Fast sweeping algorithms for a class of hamilton-jacobi equations," *SIAM Journal on Numerical Analysis*, vol. 41, no. 2, pp. 673–694, 2004.

[20] W.-K. Jeong and R. T. Whitaker, "A fast iterative method for eikonal equations," *SIAM Journal on Scientific Computing*, vol. 30, no. 5, pp. 2512–2534, Jan. 2008.

[21] N. Rawlinson and M. Sambridge, "Seismic Traveltime Tomography of the Crust and Lithosphere," *Advances in Geophysics*, vol. 46, pp. 81–198, 2003.

[22] B.-S. Shin, L. Wientgens, and D. Shutin, "Parallel 2D seismic ray tracing using CUDA on a Jetson Nano," in *ICASSP*, Jun. 2023.

[23] R. E. 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.

[24] S. Leung and J. Qian, "An adjoint state method for three dimensional refraction tomography," *Communications in Mathematical Sciences*, vol. 4, no. 1, pp. 249–266, 2006.

[25] H. Igel, *Computational Seismology*. Oxford University Press, 2017.

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

[27] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 155–171, May 2013.

[28] B.-S. Shin, H. Paul, and A. Dekorsy, "Distributed kernel least squares for nonlinear regression applied to sensor networks," in *European Signal Processing Conference*, 2016.

[29] P. R. Williamson, "A guide to the limits of resolution imposed by scattering in ray tomography," *Geophysics*, vol. 56, no. 2, pp. 202–207, Feb. 1991.

[30] A. H. Sayed, S. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Processing Magazine*, pp. 155–171, 2013.

[31] F. Cattivelli and A. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[32] B.-S. Shin, L. Wientgens, M. Schaab, and D. Shutin, "Near-surface seismic measurements in gravel pit, over highway tunnel and underground tubes with ground truth information as an open data set," *Sensors*, vol. 22, no. 17, 2022.

[33] M. Giaquinta and S. Hildebrandt, *Calculus of Variations I*. Springer Berlin Heidelberg, 2004.

PLACE PHOTO HERE

Ban-Sok Shin Biography text here.

PLACE PHOTO HERE

Dmitriy Shutin Biography text here.