

# Visco-acoustic full waveform seismic inversion: from a DG forward solver to a Newton-CG inverse solver

Thomas Bohlen, Mario R. Fernandez, Johannes Ernesti,  
Christian Rheinbay, Andreas Rieder, Christian Wieners

CRC Preprint 2020/4, February 2020

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



## Participating universities



Universität Stuttgart

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



Funded by

**DFG**

# VISCO-ACOUSTIC FULL WAVEFORM SEISMIC INVERSION: FROM A DG FORWARD SOLVER TO A NEWTON-CG INVERSE SOLVER\*

THOMAS BOHLEN<sup>†</sup>, MARIO RUBEN FERNANDEZ<sup>†</sup>, JOHANNES ERNESTI<sup>‡</sup>,  
CHRISTIAN RHEINBAY<sup>‡</sup>, ANDREAS RIEDER<sup>‡</sup>, AND CHRISTIAN WIENERS<sup>‡</sup>

**Abstract.** In this paper we present a holistic framework for full waveform inversion (FWI) in the visco-acoustic regime. FWI entails the reconstruction of material parameters (such as density and sound speed) from measurements of reflected wave fields (seismograms). We derive a discontinuous Galerkin (DG) solver for the visco-acoustic wave equation and incorporate it into an inverse solver. For the DG discretization we provide a block diagonal preconditioner for the efficient computation of the time steps by GMRES which yields a convergence estimate in space and time. Numerical tests illustrate these results. Furthermore, we set up an inverse solver of well established Newton-CG type, and we express the required Fréchet derivative and its adjoint in the DG setting. Reconstructions from simulated cross-well seismograms highlight the challenges of FWI and demonstrate the performance of the scheme. Some of the inversion experiments use seismograms generated by an independent FDTD forward solver to avoid an inverse crime.

**Key words.** full waveform inversion, Newton-CG, CG-REGINN, discontinuous Galerkin discretization, inverse crime

**AMS subject classifications.** 65M32, 65M60

**1. Introduction.** The aim of full waveform inversion (FWI) is to find material parameters such as density and wave speeds which can explain the full information content of seismic recordings. Thereby, the full signal content is fitted iteratively by full numerical solutions of the wave equation. The most important practical advantage of FWI has been discovered in the last three decades and verified in many practical applications. This is the much increased resolution of reconstructed seismic wave velocity models compared to traveltime tomography reconstructions. The second advantage, which is not yet fully exploited, is the inversion of secondary material parameter models such as seismic wave attenuation. In order to improve the reconstruction of multi-parameter models much attention has been given in recent years to derive different formulations of adjoint equations and to implement higher order optimization schemes for FWI in attenuating media. In recent years these developments have been conducted mainly in the time-domain because of the higher computational efficiency of time-domain versus frequency-domain implementations [1, 9, 27, 28].

In this work we present and combine three novel aspects of time-domain FWI in visco-acoustic media. Firstly, we present a modified formulation of the system of first order visco-acoustic wave equations which is physically more intuitive than previous formulations. This formulation, which is based on results of [29], allows the elegant derivations of adjoint equations which only show slight differences to the modified forward equations. Secondly, we implement an inexact Newton-CG method to explore its potential for the reconstruction of seismic velocity and attenuation. Thirdly, we solve the new systems of forward and adjoint equations using a higher-order discontinuous Galerkin (DG) finite-element method showing sufficient accuracy in space and time. To prove our new concept we present a first successful mono-parameter reconstruction of local anomalies of seismic wave velocity and

---

\*Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 258734477 – SFB 1173.

<sup>†</sup>Geophysical Institute, Karlsruhe Institute of Technology (KIT), D-76187 Karlsruhe, Germany (thomas.bohlen@kit.edu, mario.fernandez@kit.edu).

<sup>‡</sup>Department of Mathematics, Karlsruhe Institute of Technology (KIT), D-76128 Karlsruhe, Germany (johannes.ernesti@kit.edu, christian.rheinbay@kit.edu, andreas.rieder@kit.edu, christian.wieners@kit.edu).

attenuation in a cross-well setting. In this simplified scenario the ground truth of velocity and attenuation could efficiently be recovered even when the observed data was computed by an independent Finite Difference Time Domain (FDTD) forward solver (thereby avoiding an inverse crime).

Our paper is organized as follows. In the next section we introduce our visco-acoustic model explaining carefully the used damping mechanisms by spring and dashpot elements. Further, existence and uniqueness of the solution of the resulting evolution equation is shown in a semigroup setting. Thus, the full waveform forward operator  $\mathcal{F}$  is well defined, which maps the parameters of the visco-acoustic wave equation (density, pressure wave speed, attenuation) to the wave field. The traditional modus operandi to solve the corresponding inverse problem is the minimization of misfit functionals. We dwell a little on this approach before we introduce our algorithm **CG-REGINN** [25] of inexact Newton type which directly attacks the nonlinear operator equation. As the Fréchet derivative  $\mathcal{F}'$  and its adjoint operator are crucial ingredients we derive analytic expressions for both and explain how we evaluate those in a discrete setting. **Section 3** is devoted to our discontinuous Galerkin discretization of the visco-acoustic equation, especially to the derivation of a preconditioner for the GMRES iteration with which we compute the implicit time steps. Various convergence tests compiled in **section 4** exhibit the expected rates in time and space. Finally, the performance of our inverse solver is demonstrated elaborately by simulated cross-well experiments (**section 5**). The paper ends with a conclusion including an outlook on further research.

## 2. Full waveform inversion in the visco-acoustic regime.

**2.1. The equations of visco-acoustics for generalized standard linear solids.** The description of wave propagation in the time domain requires a rheological rock model for attenuation. For this purpose the so-called Generalized Standard Linear Solid (GSLs) is widely applied [20]. In this model  $L$  Maxwell bodies (spring  $\kappa_l$  and dashpot  $n_l$  in series;  $l = 1, \dots, L$ ) are connected in parallel with a spring  $\kappa_0$ . A combination of the single spring with one Maxwell body is called relaxation element.

The system of first order wave equations for visco-acoustic media describes the evolution of pressure  $p$  and particle velocity vector  $\mathbf{v}$ . It consists of the balance of momentum

$$(2.1a) \quad \rho \partial_t \mathbf{v} = \nabla p + \mathbf{f}$$

and the constitutive equation

$$(2.1b) \quad \partial_t p(t) = \kappa \nabla \cdot \mathbf{v}(t) + \int_0^t \dot{\kappa}(t-s) \nabla \cdot \mathbf{v}(s) ds$$

where the total unrelaxed P-wave modulus of the GSLs is  $\kappa = \kappa_0 + \dots + \kappa_L$ . The individual  $\kappa_l$  ( $l = 1, \dots, L$ ) denote the unrelaxed P-wave modulus for the  $i$ -th Maxwell body. Here,  $\kappa_0$  is the modulus of the single spring which equals the relaxed P-wave modulus of the overall GSLs.

The time derivative of the relaxation function of the GSLs reads

$$(2.1c) \quad \dot{\kappa}(s) = - \sum_{l=1}^L \frac{\kappa_l}{\tau_l} \exp\left(-\frac{s}{\tau_l}\right)$$

depending on  $L$  (unknown) unrelaxed P-wave moduli  $\kappa_l > 0$  and  $L$  (unknown) stress relaxation times  $\tau_l > 0$ ,  $l = 1, \dots, L$ . The unrelaxed P-wave moduli  $\kappa_l$  are identical and can be derived by scaling the unrelaxed P-wave modulus of the single spring:

$$(2.2a) \quad \kappa_1 = \dots = \kappa_L = \kappa_0 \tau_P.$$

In seismic and seismological applications it can be assumed that attenuation is constant within a narrow frequency range  $[\omega_{\min}, \omega_{\max}]$ . This effect is quantified by the quality factor  $Q$  which is the rate of the full energy over the dissipated energy [2]:

$$(2.2b) \quad Q(\omega) = \frac{1 + \tau_p \alpha_1(\omega)}{\alpha_2(\omega)}$$

with

$$\alpha_1(\omega) = \sum_{l=1}^L \frac{\omega^2 \tau_l^2}{1 + \omega^2 \tau_l^2} \quad \text{and} \quad \alpha_2(\omega) = \sum_{l=1}^L \frac{\omega \tau_l}{1 + \omega^2 \tau_l^2}.$$

The medium-dependent scaling factor  $\tau_p$  quantifies the level of attenuation. To realize a frequency independent attenuation ( $Q(\omega) = \text{const}$ ,  $\omega \in [\omega_{\min}, \omega_{\max}]$ ) the parameter  $\tau_p$  together with the  $L$  relaxation times  $\tau_l$  can be determined by a least-squares optimization for an assumed background medium [2, 4]. In this optimization the same relaxation times can be used for different levels of attenuation, i.e. different values of  $\tau_p$ . They can thus be determined a priori and kept fixed during the inversion process. Therefore, only the parameter  $\tau_p$  which quantifies the effects of attenuation needs to be updated.

In addition to the attenuation level  $\tau_p$  we require mass density  $\varrho$  and the relaxed P-wave modulus  $\kappa_0$  to describe waves in visco-acoustic media. By defining

$$(2.2c) \quad \kappa_0 = \frac{v_p^2 \varrho}{1 + \alpha_1(\omega_0) \tau_p}$$

we can assure that waves propagate with phase velocity  $v_p$  at a reference frequency  $\omega_0 \in [\omega_{\min}, \omega_{\max}]$ , see, e.g., [3, 4, 9].

Within this setting, the aim of visco-acoustic full waveform inversion is to identify the three parameters  $\varrho$ ,  $v_p$ , and  $\tau_p$  condensed in a parameter vector

$$\mu = (\varrho, v_p, \tau_p) \in \mathbf{P}_{\text{adm}} \subset \mathbf{P} := L_\infty(\Omega; \mathbb{R}^3)$$

within a physically meaningful parameter set

$$\mathbf{P}_{\text{adm}} = \{ \mu \in \mathbf{P} : \mu(x) \in [\varrho_{\min}, \varrho_{\max}] \times [v_{\min}, v_{\max}] \times [\tau_{\min}, \tau_{\max}] \text{ f.a.a. } x \in \Omega \}$$

for given  $0 < \varrho_{\min} < \varrho_{\max}$ ,  $0 < v_{\min} < v_{\max}$ , and  $0 < \tau_{\min} < \tau_{\max}$ .

**2.2. A semigroup setting for visco-acoustics.** The framework of monotone operators applies to (2.1) and yields uniqueness and existence of a solution as we demonstrate in this section (see also [29, 17] for visco-elasticity).

We consider a bounded Lipschitz domain  $\Omega \subset \mathbb{R}^d$  with a disjoint decomposition of the boundary  $\Gamma_{\text{dyn}} \cup \Gamma_{\text{stat}} = \partial\Omega$ , and a finite time interval  $[0, T]$ .

Introducing the decomposition of the total pressure  $p = p_0 + p_1 + \dots + p_L$  with

$$(2.3) \quad p_l(t) = \kappa_l \int_0^t \exp(-\tau_l^{-1}(t-s)) \nabla \cdot \mathbf{v}(s) \, ds, \quad l = 1, \dots, L,$$

rewrites (2.1) into the system

$$(2.4a) \quad \varrho \partial_t \mathbf{v} - \nabla p = \mathbf{f},$$

$$(2.4b) \quad \partial_t p_0 - \kappa_0 \nabla \cdot \mathbf{v} = g_0,$$

$$(2.4c) \quad \partial_t p_l - \kappa_l \nabla \cdot \mathbf{v} + \tau_l^{-1} p_l = g_l, \quad l = 1, \dots, L,$$

where we include right-hand sides in all components. It is worth noting that this formulation is equivalent to the conventional formulations where so-called memory

variables and corresponding PDEs are introduced to avoid the explicit calculation of convolutions in the constitutive equations (2.3). In our modified formulation (2.4), which is due to Zeltmann [29], we, however, avoid the introduction of memory variables. Instead, we introduce individual pressures  $p_0$  and  $p_l$  corresponding to the single spring and each Maxwell body, respectively. The overall constitutive equations then naturally separates into the constitutive relation for the spring  $(p_0, \kappa_0)$  and each Maxwell body  $(p_l, \kappa_l, \tau_l)$  where the relaxation frequencies  $\omega_l = \tau_l^{-1}$  define the center frequency at which the  $l$ -th Maxwell body operates. We therefore consider our modified system to be physically more intuitive than previous formulations.

To write (2.4) in compact operator notation we form the tuple

$$\mathbf{y} := (\mathbf{v}, p_0, \dots, p_L) \in \mathbf{Y} := L_2(\Omega; \mathbb{R}^{d+1+L}).$$

Depending on  $\mu \in \mathbf{P}_{\text{adm}}$  and the relations (2.2), we define the operators

$$(2.5a) \quad \mathbf{M}(\mu)\mathbf{y} = (\varrho\mathbf{v}, \kappa_0^{-1}p_0, \dots, \kappa_L^{-1}p_L),$$

$$(2.5b) \quad \mathbf{D}(\mu)\mathbf{y} = (\mathbf{0}, 0, (\kappa_1\tau_1)^{-1}p_1, \dots, (\kappa_L\tau_L)^{-1}p_L),$$

$$(2.5c) \quad \mathbf{A}\mathbf{y} = -(\nabla(p_0 + \dots + p_L), \text{div } \mathbf{v}, \dots, \text{div } \mathbf{v}),$$

with  $\kappa_l = \kappa_l(\mu)$  defined by (2.2). We set

$$\mathbf{L}(\mu) = \mathbf{M}(\mu)\partial_t + \mathbf{D}(\mu) + \mathbf{A}.$$

For the operator  $\mathbf{A}$  we choose a domain

$$\mathcal{D}(\mathbf{A}) = \{\mathbf{y} \in \mathbf{Y} : \mathbf{A}\mathbf{y} \in \mathbf{Y}, \mathbf{n} \cdot \mathbf{v}|_{\Gamma_{\text{dyn}}} = 0, (p_0 + \dots + p_L)|_{\Gamma_{\text{stat}}} = 0\}$$

so that  $(\mathbf{A}\mathbf{y}, \mathbf{y})_{0,\Omega} = 0$  for  $\mathbf{y} \in \mathcal{D}(\mathbf{A})$ .

Now we show that a solution  $\mathbf{y}$  of (2.4) exists, i.e., we solve

$$\mathbf{L}(\mu)\mathbf{y}^{\text{sol}}(t) = \mathbf{b}(t) \quad \text{a.e. in } (0, T), \quad \mathbf{y}(0) = \mathbf{0}.$$

LEMMA 2.1. *For given  $\mathbf{b} \in H^1(0, T; \mathbf{Y})$  the evolution problem  $\mathbf{L}(\mu)\mathbf{y}^{\text{sol}} = \mathbf{b}$  with  $\mathbf{y}^{\text{sol}}(0) = \mathbf{0}$  has a unique solution*

$$(2.6) \quad \mathbf{y}^{\text{sol}}(t) = \int_0^t \exp((s-t)\mathbf{M}(\mu)^{-1}(\mathbf{D}(\mu) + \mathbf{A}))\mathbf{M}(\mu)^{-1}\mathbf{b}(s) ds, \quad t \in [0, T],$$

in  $\mathbf{U} = \{\mathbf{y} \in H^1(0, T; \mathbf{Y}) \cap L_2(0, T; \mathcal{D}(\mathbf{A})) : \mathbf{y}(0) = \mathbf{0}\}$ .

*Remark 2.2.* For the analysis of the inverse problem in Banach spaces, regularity of the right-hand side is required, see [16, 29]. A mild solution is also obtained under weaker regularity.

*Proof.* The operator  $\mathbf{D}(\mu) + \mathbf{A}$  is monotone satisfying  $((\mathbf{D}(\mu) + \mathbf{A})\mathbf{y}, \mathbf{y})_{0,\Omega} \geq 0$  for  $\mathbf{y} \in \mathcal{D}(\mathbf{A})$ , so that the evolution equations is dissipative. Moreover, the operator  $\mathbf{M}(\mu) + \mathbf{D}(\mu) + \mathbf{A}$  is surjective, which is shown as follows:

For given  $\mathbf{g} = (\mathbf{g}_v, g_0, \dots, g_L) \in \mathbf{Y}$  consider

$$((\mathbf{M}(\mu) + \mathbf{D}(\mu) + \mathbf{A})\mathbf{y}, \mathbf{z})_{0,\Omega} = (\mathbf{g}, \mathbf{z})_{0,\Omega}, \quad \mathbf{z} \in \mathcal{D}(\mathbf{A}),$$

i.e., we construct a solution  $(\mathbf{v}, p_0, \dots, p_L) \in \mathcal{D}(\mathbf{A})$  with  $p = p_0 + \dots + p_L$  of

$$(2.7a) \quad \varrho\mathbf{v} - \nabla p = \mathbf{g}_v$$

$$(2.7b) \quad \kappa_0^{-1}p_0 - \nabla \cdot \mathbf{v} = g_0,$$

$$(2.7c) \quad \kappa_l^{-1}p_l + (\kappa_l\tau_l)^{-1}p_l - \nabla \cdot \mathbf{v} = g_l, \quad l = 1, \dots, L.$$

In the first step, we show that the solution of (2.7) in  $\mathcal{D}(\mathbf{A})$  can be obtained by solving the elliptic problem (2.9) below for  $p$  in  $V = \{q \in \mathbb{H}^1(\Omega) : q|_{\Gamma_{\text{stat}}} = 0\}$ . The following formal calculations are provided to establish (2.9). To this end, (2.7) is rescaled to

$$\begin{aligned} (2.8a) \quad & \mathbf{v} - \varrho^{-1} \nabla p = \varrho^{-1} \mathbf{g}_v \\ (2.8b) \quad & p_0 - \kappa_0 \nabla \cdot \mathbf{v} = \kappa_0 g_0, \\ (2.8c) \quad & p_l - \frac{\kappa_l \tau_l}{\tau_l + 1} \nabla \cdot \mathbf{v} = \frac{\kappa_l \tau_l}{\tau_l + 1} g_l, \quad l = 1, \dots, L, \end{aligned}$$

so that we obtain

$$\begin{aligned} p &= p_0 + \dots + p_L \\ &= \left( \kappa_0 + \frac{\kappa_1 \tau_1}{\tau_1 + 1} + \dots + \frac{\kappa_L \tau_L}{\tau_L + 1} \right) \nabla \cdot \mathbf{v} + \kappa_0 g_0 + \frac{\kappa_1 \tau_1}{\tau_1 + 1} g_1 + \dots + \frac{\kappa_L \tau_L}{\tau_L + 1} g_L \\ &= \bar{\kappa} \nabla \cdot \mathbf{v} + \bar{g} \end{aligned}$$

with  $\bar{\kappa} = \kappa_0 + \frac{\kappa_1 \tau_1}{\tau_1 + 1} + \dots + \frac{\kappa_L \tau_L}{\tau_L + 1}$  and  $\bar{g} = \kappa_0 g_0 + \frac{\kappa_1 \tau_1}{\tau_1 + 1} g_1 + \dots + \frac{\kappa_L \tau_L}{\tau_L + 1} g_L$ . Testing (2.8a) with  $q \in V$  yields

$$(\mathbf{v}, \nabla q)_{0, \Omega} - (\varrho^{-1} \nabla p, \nabla q)_{0, \Omega} = (\varrho^{-1} \mathbf{g}_v, \nabla q)_{0, \Omega}$$

and using  $\mathbf{n} \cdot \mathbf{v}|_{\Gamma_{\text{dyn}}} = 0$  gives

$$(\varrho^{-1} \nabla p, \nabla q)_{0, \Omega} = (\mathbf{v}, \nabla q)_{0, \Omega} - (\varrho^{-1} \mathbf{g}_v, \nabla q)_{0, \Omega} = -(\nabla \cdot \mathbf{v}, q)_{0, \Omega} - (\varrho^{-1} \mathbf{g}_v, \nabla q)_{0, \Omega}$$

resulting in

$$(2.9) \quad (\varrho^{-1} \nabla p, \nabla q)_{0, \Omega} + (\bar{\kappa}^{-1} p, q)_{0, \Omega} = (\bar{\kappa}^{-1} \bar{g}, q)_{0, \Omega} - (\varrho^{-1} \mathbf{g}_v, \nabla q)_{0, \Omega}, \quad q \in V.$$

Thus, we can solve (2.9) for  $p \in V$  and then we set

$$\mathbf{v} = \varrho^{-1} \nabla p + \varrho^{-1} \mathbf{g}_v,$$

so that

$$(\mathbf{v}, \nabla q)_{0, \Omega} = (\bar{\kappa}^{-1} (\bar{g} - p), q)_{0, \Omega}, \quad q \in V,$$

gives  $\mathbf{v} \in \mathbb{H}(\text{div}, \Omega)$  and  $\mathbf{n} \cdot \mathbf{v}|_{\Gamma_{\text{dyn}}} = 0$ . Defining  $p_0, \dots, p_L$  by (2.8b) and (2.8c) finally yields a solution  $(\mathbf{v}, p_0, \dots, p_L) \in \mathcal{D}(\mathbf{A})$  of (2.7).

Thus, the operator  $-\mathbf{M}(\mu)^{-1}(\mathbf{D}(\mu) + \mathbf{A})$  generates a semigroup which follows, e.g., from [23, Thm. 12.22] and the unique solution (2.6) has the stated regularity, see, e.g., [23, Thm. 12.16].  $\square$

Lemma 2.1 transfers to the adjoint problem (wave equation ‘backwards in time’)

$$(2.10) \quad \mathbf{L}(\mu)^{\text{ad}} \mathbf{y}^{\text{ad}}(t) = \mathbf{r}(t) \quad \text{a.e. in } (0, T), \quad \mathbf{y}^{\text{ad}}(T) = \mathbf{0},$$

where  $\mathbf{L}(\mu)^{\text{ad}} = -\mathbf{M}(\mu) \partial_t + \mathbf{D}(\mu) - \mathbf{A}$  is the adjoint operator such that

$$(\mathbf{L}(\mu) \mathbf{y}^{\text{sol}}, \mathbf{y}^{\text{ad}})_{0, (0, T) \times \Omega} = (\mathbf{y}^{\text{sol}}, \mathbf{L}(\mu)^{\text{ad}} \mathbf{y}^{\text{ad}})_{0, (0, T) \times \Omega}$$

for all  $\mathbf{y}^{\text{sol}}, \mathbf{y}^{\text{ad}} \in \mathbb{H}^1(0, T; \mathbf{Y}) \cap \mathbb{L}_2(0, T; \mathcal{D}(\mathbf{A}))$  with  $\mathbf{y}^{\text{sol}}(0) = \mathbf{y}^{\text{ad}}(T) = \mathbf{0}$ . Analogously to (2.6), the solution for the adjoint problem is given by

$$\mathbf{y}^{\text{ad}}(t) = \int_T^t \exp((s-t)\mathbf{M}(\mu)^{-1}(\mathbf{D}(\mu) - \mathbf{A})) \mathbf{M}(\mu)^{-1} \mathbf{b}^{\text{ad}}(s) ds, \quad t \in [0, T],$$

in  $\mathbf{U}^{\text{ad}} = \{\mathbf{y} \in \mathbb{H}^1(0, T; \mathbf{Y}) \cap \mathbb{L}_2(0, T; \mathcal{D}(\mathbf{A})) : \mathbf{y}(T) = \mathbf{0}\}$ .

**2.3. Full waveform inversion: the inverse problem.** We define the non-linear forward operator (parameter-to-solution map) by

$$\mathcal{F}: \mathbf{P}_{\text{adm}} \subset \mathbf{P} \longrightarrow \mathbf{U}, \quad \mu \longmapsto \mathbf{y},$$

i.e.,  $\mathbf{L}(\mu)\mathcal{F}(\mu) = \mathbf{b}$ .

The operator  $\mathcal{F}$  has been intensively studied in [16, 29], especially its Fréchet derivative and the adjoint operator thereof have been rigorously established in a functional analytic framework. For the reader's convenience we derive these operators below using a rather formal computation.

Let  $\mathbf{s}_{\text{obs}} \in \mathbf{S} = \mathbf{L}_2(0, T; \mathbb{R}^S)$  be a vector of measurements of the pressure at receiver points  $r_1, \dots, r_S \in \Omega$  (seismograms), and let  $\Psi: \mathbf{Y} \longrightarrow \mathbb{R}^S$  be a linear measurement operator (approximating the point evaluation). The operator extends to  $\Psi \in \mathcal{L}(\mathbf{L}_2(0, T; \mathbf{Y}), \mathbf{S})$ .

The inverse problem aims to reconstruct  $\mu \in \mathbf{P}_{\text{adm}}$  such that

$$\Psi[\mathcal{F}(\mu)] = \mathbf{s}_{\text{obs}}.$$

The above operator equation is a restricted version of a locally ill-posed problem, see [17], so that regularization is required.

In the following, we present two approaches to approximate a solution of this problem. The standard approach minimizes misfit functionals, e.g., of the form

$$(2.11) \quad \mathcal{J}(\mu) = \frac{1}{2} \|\Psi[\mathcal{F}(\mu)] - \mathbf{s}_{\text{obs}}\|_{0, (0, T)}^2$$

where  $\|\cdot\|_{0, (0, T)}$  denotes the norm in  $\mathbf{L}_2(0, T; \mathbf{Y})$ . The minimizer is obtained by computing a critical point of  $\mathcal{J}$  where the linearization  $\mathcal{J}'(\mu) \in \mathcal{L}(\mathbf{P}, \mathbb{R})$  is given by

$$(2.12) \quad \mathcal{J}'(\mu)[\delta\mu] = (\Psi[\mathcal{F}(\mu)] - \mathbf{s}_{\text{obs}}, \Psi[\mathcal{F}'(\mu)[\delta\mu]])_{0, (0, T)}$$

depending on the linearized forward operator  $\mathcal{F}'(\mu) \in \mathcal{L}(\mathbf{P}, \mathbf{U})$ .

LEMMA 2.3. *The Fréchet derivative  $\mathbf{y}^{\text{lin}} = \mathcal{F}'(\mu)[\delta\mu] \in \mathbf{U}$  is the solution of*

$$(2.13) \quad \mathbf{L}(\mu)\mathbf{y}^{\text{lin}} = -\mathbf{M}'(\mu)[\delta\mu]\partial_t \mathbf{y}^{\text{sol}} - \mathbf{D}'(\mu)[\delta\mu]\mathbf{y}^{\text{sol}}, \quad \mathbf{y}(0) = \mathbf{0},$$

where  $\mathbf{y}^{\text{sol}} = \mathcal{F}(\mu)$  solves the forward problem.<sup>1</sup>

*Proof.* Defining the Lagrange functional

$$\mathcal{L}(\mu, \mathbf{y}, \mathbf{z}) = (\mathbf{L}(\mu)\mathbf{y} - \mathbf{b}, \mathbf{z})_{0, (0, T) \times \Omega}$$

we observe that  $\mathcal{L}(\mu, \mathcal{F}(\mu), \mathbf{z}) = 0$  for all  $\mu \in \mathbf{P}_{\text{adm}}$  and  $\mathbf{z} \in \mathbf{L}_2(0, T; \mathbf{Y})$ . Taking the formal Fréchet derivative  $\mathbf{D}_\mu$  with respect to  $\mu$  yields

$$\begin{aligned} 0 &= \mathbf{D}_\mu \left( \mathcal{L}(\mu, \mathcal{F}(\mu), \mathbf{z}) \right) [\delta\mu] \\ &= \left( \mathbf{D}_\mu (\mathbf{L}(\mu)\mathcal{F}(\mu)) [\delta\mu], \mathbf{z} \right)_{0, (0, T) \times \Omega} \\ &= \left( \mathbf{M}'(\mu)[\delta\mu]\partial_t \mathbf{y}^{\text{sol}} + \mathbf{D}'(\mu)[\delta\mu]\partial_t \mathbf{y}^{\text{sol}} + \mathbf{L}(\mu)\mathcal{F}'(\mu)[\delta\mu], \mathbf{z} \right)_{0, (0, T) \times \Omega} \end{aligned}$$

and thus the assertion.  $\square$

A minimum of the misfit functional can be approximated, e.g., by a Gauß-Newton algorithm. Since the problem is ill-posed,  $\mathcal{J}$  needs to be augmented by additional regularization terms. In our approach, which we present in the following subsection, regularization is guaranteed by an inexact Newton iteration applied directly to the nonlinear problem.

<sup>1</sup>In (2.13),  $\mathbf{D}'(\mu)$  denotes the formal Fréchet derivative of the map  $\mathbf{D}: \mathbf{P}_{\text{adm}} \subset \mathbf{P} \longrightarrow \mathcal{L}(\mathbf{Y})$ ,  $\mu \longmapsto \mathbf{D}(\mu)$ , see (2.5b). Thus,  $\mathbf{D}'(\mu)[\delta\mu] \in \mathcal{L}(\mathbf{Y})$ . Same comment applies to  $\mathbf{M}$ .



**2.4. CG-REGINN: An inexact Newton-CG method.** Alternatively to the optimization problem (2.11), we can construct an iterative method solving the nonlinear equation

$$(2.14) \quad \Phi(\mu^{\text{sol}}) = \mathbf{s}_{\text{obs}},$$

where  $\Phi := \Psi \circ \mathcal{F}: \mathbf{P}_{\text{adm}} \rightarrow \mathbf{S}$  is the parameter-to-seismogram map. The realization of the Newton method requires the evaluation of the linearization  $\Phi'(\mu) \in \mathcal{L}(\mathbf{P}, \mathbf{S})$  and an appropriate initial guess  $\mu^0 \in \mathbf{P}_{\text{adm}}$ . Then, for  $k = 1, 2, \dots$ , the update  $\Delta\mu^k \in \mathbf{P}$  is determined by approximately solving  $\Phi'(\mu^k)[\Delta\mu^k] = -\mathbf{r}^k$  for  $\mathbf{r}^k = \Phi(\mu^k) - \mathbf{s}_{\text{obs}}$ . As mentioned above, the original nonlinear equation is locally ill-posed and thus we expect its linearization to be ill-posed as well. Therefore, we solve the corresponding least squares ansatz by minimizing

$$(2.15) \quad \Delta\mu^k \mapsto \frac{1}{2} \|\Phi'(\mu^k)[\Delta\mu^k] + \mathbf{r}^k\|_{0,(0,T)}^2$$

using a regularization scheme.

*Remark 2.4.* Actually, before we apply the Newton scheme we compose  $\mathcal{F}$  with a transformation yielding virtually an unconstrained equation for the transformed parameters, see [Appendix A](#). For the ease of presentation, we will stick to the formulation with constrained parameters however.

A minimizer  $\Delta\mu^k \in \mathbf{P}$  of (2.15) fulfills the normal equation

$$(2.16) \quad (\Phi'(\mu^k)[\Delta\mu^k] + \mathbf{r}^k, \Phi'(\mu^k)[\delta\mu])_{0,(0,T) \times \Omega} = 0, \quad \delta\mu \in \mathbf{P}.$$

Introducing the adjoint operator of  $\Phi'(\mu^k)$  yields a linear system in the Banach space  $\mathbf{P}'$ . Here, we solve the normal equation approximately in a discrete space  $\mathbf{P}_h \subset \mathbf{P}$  which allows to define the  $L_2$  adjoint  $\Phi'(\mu)^{\text{ad}} \in \mathcal{L}(\mathbf{S}, \mathbf{P}_h)$

$$(2.17) \quad (\Phi'(\mu)[\delta\mu], \mathbf{s})_{0,(0,T)} = (\delta\mu, \Phi'(\mu)^{\text{ad}} \mathbf{s})_{0,\Omega}, \quad \mu \in \mathbf{P}_{\text{adm}}, \delta\mu \in \mathbf{P}_h, \mathbf{s} \in \mathbf{S}.$$

Then, we solve (2.16) approximately in the discrete Hilbert space  $\mathbf{P}_h \subset L_2(\Omega; \mathbb{R}^3)$  using conjugate gradients (CG). The resulting regularization scheme for (2.14) is called CG-REGINN [25, 19], see [Algorithm 2.1](#).

---

**Algorithm 2.1** CG-REGINN for (2.14)

---

**Input:**  $\mu^0 \in \mathbf{P}_h$  % starting guess;  $\mathbf{s}_{\text{obs}} \in L_2(0, T; \mathbb{R}^S)$  % seismograms

**Output:**  $\mu^k \in \mathbf{P}_h$  % approximate solution of (2.14)

- 1:  $k \leftarrow 0, \mathbf{r}^0 \leftarrow \Phi(\mu^0) - \mathbf{s}_{\text{obs}}$
  - 2: **while not** termination **do**
  - 3: determine  $\vartheta_k$  and  $j_{\text{max},k}$  % according to (2.18) and (2.19)
  - 4:  $(\Delta\mu^k, j_k) \leftarrow \text{CG}(\mathbf{r}^k, \mu^k, \vartheta_k, j_{\text{max},k})$  % call of [Algorithm 2.2](#)
  - 5:  $\mu^{k+1} \leftarrow \mu^k + \Delta\mu^k$
  - 6:  $\mathbf{r}^{k+1} \leftarrow \Phi(\mu^{k+1}) - \mathbf{s}_{\text{obs}}$
  - 7: % Here, a safeguarding step might be added, see [Remark 2.5](#)
  - 8:  $k \leftarrow k + 1$
  - 9: **end while**
  - 10: **return**  $\mu^k$
- 

In line 4 of [Algorithm 2.1](#), the CG method ([Algorithm 2.2](#)) is called to compute the Newton update  $\Delta\mu^k$  from (2.16) in a stable way. To this end, the parameters  $\vartheta_k$  and  $j_{\text{max},k}$  are passed to the CG method. They steer adaptively the magnitude of regularization depending on the local degree of ill-posedness by stopping the

iteration either because the residual of the actual iterate is sufficiently small or a maximal number of iteration steps has been reached (line 12 of Algorithm 2.2). In line 3 of Algorithm 2.1 we used the following strategy for choosing  $\vartheta_k$  and  $j_{\max,k}$  (other strategies are also possible, see [24, Sec. 6] for an explanation):

$$(2.18) \quad \vartheta_k := \min \{0.999, \tilde{\vartheta}_k\}$$

where

$$\tilde{\vartheta}_k := \begin{cases} 1 & : k = 0, \\ \|\mathbf{r}^1\|_{0,(0,T)} / \|\mathbf{r}^0\|_{0,(0,T)} & : k = 1, \\ 1 - \frac{j_{k-2}}{j_{k-1}} (1 - \vartheta_{k-1}) & : j_{k-1} > j_{k-2} \wedge k \geq 2, \\ 0.9 \vartheta_{k-1} & : \text{otherwise.} \end{cases}$$

Finally,

$$(2.19) \quad j_{\max,k} := \begin{cases} 1 & : k = 0, \\ 2 & : k = 1, \\ j_{k-1} + j_{k-2} & : k \geq 2. \end{cases}$$

---

**Algorithm 2.2** Conjugate Gradient (CG) algorithm for (2.16) restricted to  $\mathbf{P}_h$

---

**Input:**  $\mathbf{r}^k \in \mathbf{S}$ ,  $\mu^k \in \mathbf{P}_h$ ,  $\vartheta \in (0, 1)$ ,  $j_{\max} \in \mathbb{N}$

**Output:**  $\Delta\mu^k \in \mathbf{P}_h$ ,  $j_k \in \mathbb{N}$

- 1:  $j \leftarrow 0$ ,  $\beta \leftarrow 0$ ,  $\tilde{\mathbf{r}}^0 \leftarrow \mathbf{r}^k$
  - 2:  $p^0, f^0 \leftarrow 0 \in \mathbf{P}_h$
  - 3: **repeat**
  - 4:  $j \leftarrow j + 1$
  - 5:  $d \leftarrow \Phi'(\mu^k)^{\text{ad}}[\tilde{\mathbf{r}}^{j-1}] \in \mathbf{P}_h$
  - 6:  $p^j \leftarrow d + \beta \|d\|_{0,\Omega}^2 p^{j-1}$
  - 7:  $\mathbf{q} \leftarrow \Phi'(\mu^k)[p^j] \in \mathbf{S}$
  - 8:  $\alpha \leftarrow \|d\|_{0,\Omega}^2 / \|\mathbf{q}\|_{0,(0,T)}^2$
  - 9:  $f^j \leftarrow f^{j-1} + \alpha p^j$
  - 10:  $\tilde{\mathbf{r}}^j \leftarrow \tilde{\mathbf{r}}^{j-1} - \alpha \mathbf{q}$
  - 11:  $\beta \leftarrow 1 / \|d\|_{0,\Omega}^2$
  - 12: **until**  $\|\tilde{\mathbf{r}}^j\|_{0,(0,T)} \leq \vartheta \|\mathbf{r}^k\|_{0,(0,T)}$  **or**  $j \geq j_{\max}$
  - 13:  $\Delta\mu^k \leftarrow f^j$ ,  $j_k \leftarrow j$
  - 14: **return**  $(\Delta\mu^k, j_k)$
- 

*Remark 2.5.* In a reasonable mathematical setting, the nonlinear residuals  $\mathbf{r}^k = \Phi(\mu^k) - \mathbf{s}_{\text{obs}}$  can be shown to be strictly decreasing [19, Thm 3.1]. In the geophysical applications, it is not clear whether the mathematical assumptions are justified. Therefore one should check for monotonicity and, if necessary, enforce it by back-tracking or other strategies:

```

if  $\|\mathbf{r}^{k+1}\|_{0,(0,T)} \geq \|\mathbf{r}^k\|_{0,(0,T)}$  then
   $t \leftarrow 1$ 
  repeat
     $t \leftarrow 0.9 t$ 
     $\mu^{k+1} \leftarrow \mu^k + t \Delta\mu^k$ 
     $\mathbf{r}^{k+1} \leftarrow \Phi(\mu^{k+1}) - \mathbf{s}_{\text{obs}}$ 
  until  $\|\mathbf{r}^{k+1}\|_{0,(0,T)} < \|\mathbf{r}^k\|_{0,(0,T)}$ 

```

end if

The above repeat-loop terminates as  $\Delta\mu^k$  is a decent direction at  $\mu^k$  of the misfit functional  $\mathcal{J}$  (2.11), see [19, Lem. 2.2].

**2.5. Computing the Fréchet derivative and its adjoint.** The realization of the CG method requires to evaluate  $\Phi'(\mu)[\delta\mu]$  and  $\Phi'(\mu)^{\text{ad}}\mathbf{s}$ . By definition of  $\Phi$  and the linearity of  $\Psi$ , we have

$$\Phi'(\mu) = \Psi \circ \mathcal{F}'(\mu), \quad \Phi'(\mu)^{\text{ad}} = \mathcal{F}'(\mu)^{\text{ad}} \circ \Psi^{\text{ad}},$$

where the adjoint operator  $\mathcal{F}'(\mu)^{\text{ad}} \in \mathcal{L}(\mathbf{L}_2(0, T; \mathbf{Y}), \mathbf{P}_h)$  is given by

$$(\mathcal{F}'(\mu)[\delta\mu], \mathbf{z})_{0, (0, T) \times \Omega} = (\delta\mu, \mathcal{F}'(\mu)^{\text{ad}}[\mathbf{z}])_{0, \Omega}, \quad \delta\mu \in \mathbf{P}_h, \quad \mathbf{z} \in \mathbf{L}_2(0, T; \mathbf{Y}).$$

LEMMA 2.6. *We have that  $\mu^{\text{ad}} = \mathcal{F}'(\mu)^{\text{ad}}[\mathbf{z}] \in \mathbf{P}_h$  is given as the solution of the variational problem*

$$(\mu^{\text{ad}}, \delta\mu)_{0, \Omega} = -(\mathbf{M}'(\mu)[\delta\mu]\partial_t \mathbf{y}^{\text{sol}} + \mathbf{D}'(\mu)[\delta\mu]\mathbf{y}^{\text{sol}}, \mathbf{y}^{\text{ad}})_{0, (0, T) \times \Omega}, \quad \delta\mu \in \mathbf{P}_h,$$

where  $\mathbf{y}^{\text{sol}} = \mathcal{F}(\mu)$  solves the forward problem and  $\mathbf{y}^{\text{ad}} \in \mathbf{U}^{\text{ad}}$  solves

$$(2.20) \quad \mathbf{L}(\mu)^{\text{ad}}\mathbf{y}^{\text{ad}} = \mathbf{z}, \quad \mathbf{y}^{\text{ad}}(T) = 0,$$

see (2.10).

*Proof.* The assertion follows readily from  $\mathbf{y}^{\text{lin}} = \mathcal{F}'(\mu)[\delta\mu]$  and

$$\begin{aligned} (\mathcal{F}'(\mu)^{\text{ad}}[\mathbf{z}], \delta\mu)_{0, \Omega} &= (\mathbf{z}, \mathcal{F}'(\mu)[\delta\mu])_{0, (0, T) \times \Omega} \\ &= (\mathbf{L}(\mu)^{\text{ad}}\mathbf{y}^{\text{ad}}, \mathbf{y}^{\text{lin}})_{0, (0, T) \times \Omega} \\ &= (\mathbf{y}^{\text{ad}}, \mathbf{L}(\mu)\mathbf{y}^{\text{lin}})_{0, (0, T) \times \Omega} \\ &= -(\mathbf{y}^{\text{ad}}, \mathbf{M}'(\mu)[\delta\mu]\partial_t \mathbf{y}^{\text{sol}} + \mathbf{D}'(\mu)[\delta\mu]\mathbf{y}^{\text{sol}})_{0, (0, T) \times \Omega} \end{aligned}$$

using (2.20) and (2.13).  $\square$

In our application, the discrete parameter space  $\mathbf{P}_h$  contains functions being cell-wise constant with respect to the mesh  $\bar{\Omega} = \bigcup_{K \in \mathcal{K}} \bar{K}$  (see the following section for more details about the mesh). Thus, we obtain from Lemma 2.6 by testing with basis vectors  $\{\mathbf{e}_K^j : K \in \mathcal{K}, j = 1, 2, 3\}$  of  $\mathbf{P}_h$  that

$$\mu_j^{\text{ad}}|_K = \frac{-1}{|K|} (\partial_{\mu_j} \mathbf{M}(\mu)|_K \partial_t \mathbf{y}^{\text{sol}} + \partial_{\mu_j} \mathbf{D}(\mu)|_K \mathbf{y}^{\text{sol}}, \mathbf{y}^{\text{ad}})_{0, (0, T) \times K}$$

with the derivatives  $\partial_{\mu_j} \mathbf{M}(\mu)|_K = \mathbf{M}'(\mu)[\mathbf{e}_K^j]$  and  $\partial_{\mu_j} \mathbf{D}(\mu)|_K = \mathbf{D}'(\mu)[\mathbf{e}_K^j]$ .

By the definitions of  $\kappa_0 = \frac{v_p^2 \varrho}{1 + \alpha \tau_p}$  with  $\alpha = \alpha_1(\omega_0)$  and  $\kappa_l = \tau_p \kappa_0$ ,  $l = 1, \dots, L$ , see (2.2), we obtain

$$\begin{aligned} \partial_{\varrho} \kappa_l^{-1} &= -\varrho^{-1} \kappa_l^{-1}, & \partial_{v_p} \kappa_l^{-1} &= -2v_p^{-1} \kappa_l^{-1}, & l &= 0, \dots, L, \\ \partial_{\tau_p} \kappa_0^{-1} &= v_p^{-2} \varrho^{-1} \alpha, & \partial_{\tau_p} \kappa_l^{-1} &= -(v_p \tau_p)^{-2} \varrho^{-1}, & l &= 1, \dots, L. \end{aligned}$$

In view of (2.5) we thus get for  $\mathbf{y} = (\mathbf{v}, p_0, \dots, p_L) \in \mathbb{R}^{d+1+L}$

$$\begin{aligned} \partial_{\varrho} \mathbf{M}(\mu)|_K \mathbf{y} &= -\varrho_K^{-1} \kappa_{0,K}^{-1} (\mathbf{v}, p_0, \tau_{p,K}^{-1} p_1, \dots, \tau_{p,K}^{-1} p_L), \\ \partial_{v_p} \mathbf{M}(\mu)|_K \mathbf{y} &= -2v_{p,K}^{-1} \kappa_{0,K}^{-1} (\mathbf{v}, p_0, \tau_{p,K}^{-1} p_1, \dots, \tau_{p,K}^{-1} p_L), \end{aligned}$$

$$\partial_{\tau_p} \mathbf{M}(\mu)|_K \mathbf{y} = -v_{p,K}^{-2} \varrho_K^{-1} (\mathbf{0}, -\alpha p_0, \tau_{p,K}^{-2} p_1, \dots, \tau_{p,K}^{-2} p_L),$$

and

$$\partial_\varrho \mathbf{D}(\mu)|_K \mathbf{y} = -\varrho_K^{-1} \kappa_{0,K}^{-1} \tau_{p,K}^{-1} (\mathbf{0}, 0, \tau_1^{-1} p_1, \dots, \tau_L^{-1} p_L),$$

$$\partial_{v_p} \mathbf{D}(\mu)|_K \mathbf{y} = -2v_{p,K}^{-1} \kappa_{0,K}^{-1} \tau_{p,K}^{-1} (\mathbf{0}, 0, \tau_1^{-1} p_1, \dots, \tau_L^{-1} p_L),$$

$$\partial_{\tau_p} \mathbf{D}(\mu)|_K \mathbf{y} = -(v_{p,K} \tau_{p,K})^{-2} \varrho_K^{-1} (\mathbf{0}, 0, \tau_1^{-1} p_1, \dots, \tau_L^{-1} p_L).$$

For  $(\varrho_K^{\text{ad}}, v_{p,K}^{\text{ad}}, \tau_{p,K}^{\text{ad}})_{K \in \mathcal{K}} = \mathcal{F}'(\mu)^{\text{ad}}[\mathbf{b}] \in \mathbf{P}_h$  and  $\mathbf{y}^{\text{ad}} = (\mathbf{w}^{\text{ad}}, q_0^{\text{ad}}, \dots, q_L^{\text{ad}})$  we conclude for each  $K \in \mathcal{K}$  that

$$(2.21a) \quad \varrho_K^{\text{ad}} = \frac{1}{|K|} \int_0^T \int_K \left( \partial_t \mathbf{v}^{\text{sol}} \cdot \mathbf{w}^{\text{ad}} - \frac{1}{\kappa_{0,K} \varrho_K} \left( p_0^{\text{sol}} q_0^{\text{ad}} + \frac{1}{\tau_{p,K}} \sum_{l=1}^L (\partial_t p_l^{\text{sol}} + \frac{1}{\tau_l} p_l^{\text{sol}}) q_l^{\text{ad}} \right) \right) dx dt,$$

$$(2.21b) \quad v_{p,K}^{\text{ad}} = \frac{1}{|K|} \frac{-2}{\kappa_{0,K} v_{p,K}} \int_0^T \left( \int_K \partial_t p_0^{\text{sol}} q_0^{\text{ad}} + \frac{1}{\tau_{p,K}} \sum_{l=1}^L (\partial_t p_l^{\text{sol}} + \frac{1}{\tau_l} p_l^{\text{sol}}) q_l^{\text{ad}} \right) dx dt,$$

$$(2.21c) \quad \tau_{p,K}^{\text{ad}} = \frac{-1}{|K|} \int_0^T \int_K \left( \frac{\alpha}{v_p^2 \varrho_K} \partial_t p_0^{\text{sol}} q_0^{\text{ad}} + \frac{1}{v_{p,K} \tau_{p,K}^2 \varrho_K} \sum_{l=1}^L (\partial_t p_l^{\text{sol}} + \frac{1}{\tau_l} p_l^{\text{sol}}) q_l^{\text{ad}} \right) dx dt.$$

*Remark 2.7.* The above construction restricts the operators  $\Phi'(\mu)$  and  $\mathcal{F}'(\mu)$  to the discrete space  $\mathbf{P}_h$  in order to define proper adjoints in an  $L_2$  sense. The natural, infinite dimensional formulation involves the Banach space  $\mathbf{P} = L_\infty(\Omega, \mathbb{R}^3)$ , see [16, 17, 29]. By identifying  $L_2(0, T; \mathbf{Y})$  and  $L_2(0, T; \mathbb{R}^S)$  with their duals, this yields the Banach space adjoint

$$\Phi'(\mu)^* = \mathcal{F}'(\mu)^* \Psi^{\text{ad}} \text{ with } \mathcal{F}'(\mu)^* \in \mathcal{L}(L_2(0, T; \mathbf{Y}), L_\infty(\Omega, \mathbb{R}^3)').$$

Following the proof of Lemma 2.6, this implies for  $\delta\mu \in \mathbf{P}$  and  $\mathbf{z} \in L_2(0, T; \mathbf{Y})$

$$\begin{aligned} \langle \mathcal{F}'(\mu)^*[\mathbf{z}], \delta\mu \rangle_{\mathbf{P}', \mathbf{P}} &= (\mathbf{z}, \mathcal{F}'(\mu)[\delta\mu])_{0, (0, T) \times \Omega} \\ &= -(\mathbf{M}'(\mu)[\delta\mu] \partial_t \mathbf{y}^{\text{sol}} + \mathbf{D}'(\mu)[\delta\mu] \mathbf{y}^{\text{sol}}, \mathbf{y}^{\text{ad}})_{0, (0, T) \times \Omega} \\ &= \sum_{j=1}^3 \left( \delta\mu_j, \int_0^T (\partial_{\mu_j} \mathbf{M}(\mu) \partial_t \mathbf{y}^{\text{sol}} + \partial_{\mu_j} \mathbf{D}(\mu) \mathbf{y}^{\text{sol}}) \cdot \mathbf{y}^{\text{ad}} dt \right)_{0, \Omega} \end{aligned}$$

with  $\mathbf{M}'(\mu)[\delta\mu] = \sum_{j=1}^3 \delta\mu_j \partial_{\mu_j} \mathbf{M}(\mu)$  and  $\mathbf{D}'(\mu)[\delta\mu] = \sum_{j=1}^3 \delta\mu_j \partial_{\mu_j} \mathbf{D}(\mu)$ . Thus, for each component of  $(\mu_1^*, \mu_2^*, \mu_3^*) = \mathcal{F}'(\mu)^*[\mathbf{z}]$ , we obtain

$$(2.22) \quad \mu_j^* = \int_0^T (\partial_{\mu_j} \mathbf{M}(\mu) \partial_t \mathbf{y}^{\text{sol}} + \partial_{\mu_j} \mathbf{D}(\mu) \mathbf{y}^{\text{sol}}) \cdot \mathbf{y}^{\text{ad}} dt \in L_\infty(\Omega)'$$

Since  $\mathbf{y}^{\text{sol}}, \mathbf{y}^{\text{ad}} \in L_2(0, T; \mathbf{Y})$  the integral represents in fact a function in  $L_1(\Omega, \mathbb{R}^3)$  and (2.21) is obtained by averaging (2.22) for each cell  $K \in \mathcal{K}$ .

**3. A discontinuous Galerkin method.** The visco-acoustic wave equation is approximated in the discontinuous finite element space

$$\mathbf{Y}_h = \{ \mathbf{y}_h \in \mathbf{L}_2(\Omega; \mathbb{R}^{d+1+L}) : \mathbf{y}_h|_K \in \mathbb{P}_k(\mathbb{R}^{d+1+L}) \text{ for all } K \in \mathcal{K} \}$$

with polynomial degree  $k \geq 0$ . Here, the domain  $\Omega$  is decomposed into open convex polyedral sets  $K \subset \Omega$ , called *cells*, that is,  $\bar{\Omega} = \bigcup_{K \in \mathcal{K}} \bar{K}$  ( $\mathcal{K}$  is the set of all cells).

Let  $\mathcal{F}_K$  be the set of faces of  $K$ , and set  $\mathcal{F} = \bigcup_K \mathcal{F}_K$ . For inner faces  $f \in \mathcal{F} \cap \Omega$ , let  $K_f$  be the neighboring cell such that  $\bar{f} = \partial K \cap \partial K_f$ . On boundary faces  $f \in \mathcal{F} \cap \partial\Omega$  we set  $K_f = K$ . Let  $\mathbf{n}_K$  be the outer unit normal vector on  $\partial K$ .

For  $\mathbf{y}_h \in \mathbf{Y}_h$  the restriction to  $K$  is denoted by  $\mathbf{y}_{h,K} = \mathbf{y}_h|_K$ . On inner faces  $f \in \mathcal{F} \cap \Omega$  we define  $[\mathbf{y}_h]_{K,f} = \mathbf{y}_{h,K_f} - \mathbf{y}_{h,K}$ . On boundary faces  $f \in \mathcal{F} \cap \Gamma_{\text{kin}}$ , we set  $[\mathbf{v}_h]_{K,f} = \mathbf{0}$  and  $[p_h]_{K,f} = -2p_h$ , and on  $f \in \mathcal{F} \cap \Gamma_{\text{stat}}$  we set  $[\mathbf{v}_h]_{K,f} = -2\mathbf{v}_h$  and  $[p_h]_{K,f} = 0$ .

For fixed  $\mu \in \mathbf{P}_h$  and  $t \in (0, T)$ , we consider the discretized problem in space

$$(3.1) \quad \mathbf{M}_h \partial_t \mathbf{y}_h(t) + \mathbf{A}_h \mathbf{y}_h(t) + \mathbf{D}_h \mathbf{y}_h(t) = \mathbf{b}_h(t)$$

where the discrete operators  $\mathbf{M}_h, \mathbf{D}_h \in \mathcal{L}(\mathbf{Y}_h, \mathbf{Y}_h)$  and the right-hand side  $\mathbf{b}_h \in \mathbf{L}_2(0, T; \mathbf{Y}_h)$  are the Galerkin approximations defined by

$$\begin{aligned} (\mathbf{M}_h \mathbf{y}_h, \mathbf{z}_h)_{0,\Omega} &= (\mathbf{M} \mathbf{y}_h, \mathbf{z}_h)_{0,\Omega}, \\ (\mathbf{D}_h \mathbf{y}_h, \mathbf{z}_h)_{0,\Omega} &= (\mathbf{D} \mathbf{y}_h, \mathbf{z}_h)_{0,\Omega}, \\ (\mathbf{b}_h, \mathbf{z}_h)_{0,\Omega} &= (\mathbf{b}, \mathbf{z}_h)_{0,\Omega}, \quad \mathbf{y}_h, \mathbf{z}_h \in \mathbf{Y}_h. \end{aligned}$$

Further,  $\mathbf{A}_h = \sum_{K \in \mathcal{K}} \mathbf{A}_{h,K} \in \mathcal{L}(\mathbf{Y}_h, \mathbf{Y}_h)$  is the DG approximation with full upwind flux given by

$$\begin{aligned} (\mathbf{A}_{h,K} \mathbf{y}_h, \mathbf{z}_h)_{0,K} &= -(\nabla p_{h,K}, \mathbf{w}_{h,K})_{0,K} - (\nabla \cdot \mathbf{v}_{h,K}, q_{h,K})_{0,K} \\ &\quad - \sum_{f \in \mathcal{F}_K} \frac{1}{Z_K + Z_{K_f}} ([p_h]_{K,f} + Z_{K_f} \mathbf{n}_K \cdot [\mathbf{v}_h]_{K,f}, q_{K,h} + Z_K \mathbf{n}_K \cdot \mathbf{w}_{h,K})_{0,f} \end{aligned}$$

for  $\mathbf{y}_h = (\mathbf{v}_h, p_{0,h}, \dots, p_{L,h})$ ,  $\mathbf{z}_h = (\mathbf{w}_h, q_{0,h}, \dots, q_{L,h}) \in \mathbf{Y}_h$  with  $p_h = p_{0,h} + \dots + p_{L,h}$ ,  $q_h = q_{0,h} + \dots + q_{L,h}$  and impedance  $Z_K = \sqrt{\kappa \varrho}|_K$ , see [13, 14] for details. For the construction and the subsequent analysis of the DG method, we assume that the material parameters are constant in every  $K \in \mathcal{K}$ .

In time we use the implicit mid-point rule with fixed time step size  $\Delta t = T/N$  and time step  $t_n = n\Delta t$ . Starting with  $\mathbf{y}_h^0 = \mathbf{0}$ , the approximations  $\mathbf{y}_h^n \in \mathbf{Y}_h$  of (3.1) are defined by

$$\mathbf{M}_h \partial_{\Delta t/2} \mathbf{y}_h^{n-1/2} + \mathbf{A}_h \mathbf{y}_h^{n-1/2} + \mathbf{D}_h \mathbf{y}_h^{n-1/2} = \mathbf{b}_h^{n-1/2}, \quad n = 1, \dots, N,$$

with  $\partial_{\Delta t/2} \mathbf{y}_h^{n-1/2} = \frac{1}{\Delta t} (\mathbf{y}_h^n - \mathbf{y}_h^{n-1})$ ,  $\mathbf{y}_h^{n-1/2} = \frac{1}{2} (\mathbf{y}_h^{n-1} + \mathbf{y}_h^n)$ , and  $\mathbf{b}_h^{n-1/2} = \mathbf{b}_h(t_{n-1/2})$ , so that  $\mathbf{y}_h^n$  is determined by the linear system

$$(3.2) \quad \left( \mathbf{M}_h + \frac{\Delta t}{2} \mathbf{A}_h + \frac{\Delta t}{2} \mathbf{D}_h \right) \mathbf{y}_h^n = \left( \mathbf{M}_h - \frac{\Delta t}{2} \mathbf{A}_h - \frac{\Delta t}{2} \mathbf{D}_h \right) \mathbf{y}_h^{n-1} + \Delta t \mathbf{b}_h^{n-1/2}.$$

Note that this scheme is reversible and unconditionally stable.

Now we show that the linear system (3.2) is well-posed and we estimate the condition number with respect to the energy norm. To this end we introduce the energy norm  $\|\mathbf{y}_h\|_{\mathbf{M}_h}^2 = (\mathbf{y}_h, \mathbf{y}_h)_{\mathbf{M}_h}$  via the weighted inner product  $(\mathbf{y}_h, \mathbf{y}_h)_{\mathbf{M}_h} = (\mathbf{M}_h \mathbf{y}_h, \mathbf{y}_h)_{0,\Omega}$ .

LEMMA 3.1. *We have for  $\mathbf{y}_h \in \mathbf{Y}_h$  that*

$$(3.3a) \quad \left( \mathbf{M}_h^{-1} \left( \mathbf{M}_h + \frac{\Delta t}{2} \mathbf{A}_h + \frac{\Delta t}{2} \mathbf{D}_h \right) \mathbf{y}_h, \mathbf{y}_h \right)_{\mathbf{M}_h} \geq \|\mathbf{y}_h\|_{\mathbf{M}_h}^2,$$

$$(3.3b) \quad \left\| \mathbf{M}_h^{-1} \left( \mathbf{M}_h + \frac{\Delta t}{2} \mathbf{A}_h + \frac{\Delta t}{2} \mathbf{D}_h \right) \mathbf{y}_h \right\|_{\mathbf{M}_h} \leq C \|\mathbf{y}_h\|_{\mathbf{M}_h}$$

with  $C = 1 + \Delta t (C_1 v_{p,\text{ref}} h^{-1} + C_2)$  where  $C_1 \geq 0$  depends on the mesh quality, the polynomial degree  $k$  of the discrete space  $\mathbf{Y}_h$ , and on material parameters with respect to a reference velocity  $v_{p,\text{ref}} > 0$ . The constant  $C_2 \geq 0$  only depends on the relaxation parameters  $\tau_l$ ,  $l = 1, \dots, L$ .

*Proof.* We have for  $\mathbf{y}_h = (\mathbf{v}_h, p_{0,h}, \dots, p_{L,h}) \in \mathbf{Y}_h$  and  $p_h = p_{0,h} + \dots + p_{L,h}$

$$(\mathbf{A}_h \mathbf{y}_h, \mathbf{y}_h)_{0,\Omega} = \frac{1}{2} \sum_{K \in \mathcal{K}} \sum_{f \in \mathcal{F}_K} \frac{1}{Z_K + Z_{K_f}} \left( \|[p_h]_{K,f}\|_{0,f}^2 + Z_K Z_{K_f} \|\mathbf{n}_K \cdot [\mathbf{v}_h]_{K,f}\|_{0,f}^2 \right),$$

cf. [14, Sect. 4.3]. Further,  $(\mathbf{D}_h \mathbf{y}_h, \mathbf{y}_h)_{0,\Omega} \geq 0$  and

$$\left( \mathbf{M}_h^{-1} \left( \mathbf{M}_h + \frac{\Delta t}{2} \mathbf{A}_h + \frac{\Delta t}{2} \mathbf{D}_h \right) \mathbf{y}_h, \mathbf{y}_h \right)_{\mathbf{M}_h} = \|\mathbf{y}_h\|_{\mathbf{M}_h}^2 + \frac{\Delta t}{2} ((\mathbf{A}_h + \mathbf{D}_h) \mathbf{y}_h, \mathbf{y}_h)_{0,\Omega}.$$

Hence, (3.3a) holds true. The second assertion (3.3b) follows from

$$\begin{aligned} & \left\| \mathbf{M}_h^{-1} \left( \mathbf{M}_h + \frac{\Delta t}{2} \mathbf{A}_h + \frac{\Delta t}{2} \mathbf{D}_h \right) \mathbf{y}_h \right\|_{\mathbf{M}_h} \\ & \leq \|\mathbf{y}_h\|_{\mathbf{M}_h} + \frac{\Delta t}{2} \left( \|\mathbf{M}_h^{-1} \mathbf{A}_h \mathbf{y}_h\|_{\mathbf{M}_h} + \|\mathbf{M}_h^{-1} \mathbf{D}_h \mathbf{y}_h\|_{\mathbf{M}_h} \right) \end{aligned}$$

and from estimates for  $\|\mathbf{M}_h^{-1} \mathbf{A}_h \mathbf{y}_h\|_{\mathbf{M}_h}$  and  $\|\mathbf{M}_h^{-1} \mathbf{D}_h \mathbf{y}_h\|_{\mathbf{M}_h}$ . Therefore, we use the inverse inequalities [6, Chap. 1.4.3]

$$\begin{aligned} h \|\nabla p_h\|_{0,K} + h^{1/2} \|p_h\|_{0,\partial K} & \leq C_{\text{inv}} \|p_h\|_{0,K}, \\ h \|\nabla \cdot \mathbf{v}_h\|_{0,K} + h^{1/2} \|\mathbf{n} \cdot \mathbf{v}_h\|_{0,\partial K} & \leq C_{\text{inv}} \|\mathbf{v}_h\|_{0,K}, \quad (\mathbf{v}_h, p_h, 0, \dots, 0) \in \mathbf{Y}_h, \quad K \in \mathcal{K}, \end{aligned}$$

with  $C_{\text{inv}} > 0$  depending on the mesh quality and the polynomial degree  $k$  of the discrete space  $\mathbf{Y}_h$ . We introduce a reference velocity  $v_{p,\text{ref}} > 0$  such that  $\sqrt{\kappa/\rho} \leq v_{p,\text{ref}}$ . Then, there exists a constant  $C_{\text{param}} > 0$  only depending on the parameter variation such that  $\|\kappa^{-1/2} p_h\|_{0,\Omega} + \|\rho^{1/2} \mathbf{v}_h\|_{0,\Omega} \leq C_{\text{param}} \|\mathbf{y}_h\|_{\mathbf{M}_h}$ .

Now, with  $\mathbf{z}_h = \mathbf{M}_h^{-1} \mathbf{A}_h \mathbf{y}_h = (\mathbf{w}_h, q_{0,h}, \dots, q_{L,h}) \in \mathbf{Y}_h$  and  $q_h = q_{0,h} + \dots + q_{L,h}$  we obtain

$$\begin{aligned} \|\mathbf{M}_h^{-1} \mathbf{A}_h \mathbf{y}_h\|_{\mathbf{M}_h}^2 & = (\mathbf{A}_h (\mathbf{v}_h, p_{0,h}, \dots, p_{L,h}), (\mathbf{w}_h, q_{0,h}, \dots, q_{L,h}))_{0,\Omega} \\ & = -(\nabla p_h, \mathbf{w}_h)_{0,\Omega} - (\nabla \cdot \mathbf{v}_h, q_h)_{0,\Omega} \\ & \quad + \sum_K \sum_{f \in \mathcal{F}_K} \frac{1}{Z_K + Z_{K_f}} \left( [p_h]_{K,f} + Z_{K_f} \mathbf{n}_K \cdot [\mathbf{v}_h]_{K,f}, q_{K,h} + Z_K \mathbf{n}_K \cdot \mathbf{w}_{h,K} \right)_{0,f} \\ & \leq C_{\text{inv}} h^{-1} \left( \|p_h\|_{0,\Omega} \|\mathbf{w}_h\|_{0,\Omega} + \|\mathbf{v}_h\|_{0,\Omega} \|q_h\|_{0,\Omega} \right. \\ & \quad \left. + \sum_K \sum_{f \in \mathcal{F}_K} \left( \frac{1}{Z_K + Z_{K_f}} \|p_h\|_{0,K} \|q_K\|_{0,K} + \frac{Z_{K_f}}{Z_K + Z_{K_f}} \|\mathbf{v}_h\|_{0,K} \|q_h\|_{0,K} \right. \right. \\ & \quad \left. \left. + \frac{Z_K}{Z_K + Z_{K_f}} \|p_h\|_{0,K} \|\mathbf{w}_h\|_{0,K} + \frac{Z_K Z_{K_f}}{Z_K + Z_{K_f}} \|\mathbf{v}_h\|_{0,K} \|\mathbf{w}_h\|_{0,K} \right) \right) \end{aligned}$$

$$\begin{aligned}
&\leq C_{\text{inv}v_{\text{p,ref}}}h^{-1} \left( \|\kappa^{-1/2}p_h\|_{0,\Omega} \|\rho^{1/2}\mathbf{w}_h\|_{0,\Omega} + \|\rho^{1/2}\mathbf{v}_h\|_{0,\Omega} \|\kappa^{-1/2}q_h\|_{0,\Omega} \right. \\
&\quad + \sum_K \sum_{f \in \mathcal{F}_K} \left( \|\kappa^{-1/2}p_h\|_{0,K} \|\kappa^{-1/2}q_K\|_{0,K} + \|\rho^{1/2}\mathbf{v}_h\|_{0,K} \|\kappa^{-1/2}q_h\|_{0,K} \right. \\
&\quad \quad \left. \left. + \|\kappa^{-1/2}p_h\|_{0,K} \|\rho^{1/2}\mathbf{w}_h\|_{0,K} + \|\rho^{1/2}\mathbf{v}_h\|_{0,K} \|\rho^{1/2}\mathbf{w}_h\|_{0,K} \right) \right) \\
&\leq C_{\text{inv}v_{\text{p,ref}}}h^{-1} \left( (\|\kappa^{-1/2}p_h\|_{0,\Omega} + \|\rho^{1/2}\mathbf{v}_h\|_{0,\Omega}) (\|\kappa^{-1/2}q_h\|_{0,\Omega} + \|\rho^{1/2}\mathbf{w}_h\|_{0,\Omega}) \right. \\
&\quad \left. + \sum_K \sum_{f \in \mathcal{F}_K} (\|\kappa^{-1/2}p_h\|_{0,K} + \|\rho^{1/2}\mathbf{v}_h\|_{0,K}) (\|\kappa^{-1/2}q_K\|_{0,K} + \|\rho^{1/2}\mathbf{w}_h\|_{0,K}) \right) \\
&\leq 3C_{\text{inv}v_{\text{p,ref}}}h^{-1} \left( \|\kappa^{-1/2}p_h\|_{0,\Omega} + \|\rho^{1/2}\mathbf{v}_h\|_{0,\Omega} \right) \left( \|\kappa^{-1/2}q_h\|_{0,\Omega} + \|\rho^{1/2}\mathbf{w}_h\|_{0,\Omega} \right) \\
&\leq 3C_{\text{inv}v_{\text{p,ref}}}h^{-1} C_{\text{param}} \|\mathbf{y}_h\|_{M_h} \|\mathbf{z}_h\|_{M_h}
\end{aligned}$$

which yields  $\|M_h^{-1}A_h\mathbf{y}_h\|_{M_h} \leq 3C_{\text{inv}v_{\text{p,ref}}}h^{-1}C_{\text{param}}\|\mathbf{y}_h\|_{M_h}$ . Taking into account that

$$\begin{aligned}
(M_h^{-1}D_h\mathbf{y}_h, D_h\mathbf{y}_h)_{0,\Omega} &= (\kappa_1^{-1}\tau_1^{-2}p_{1,h}, p_{1,h})_{0,\Omega} + \cdots + (\kappa_L^{-1}\tau_L^{-2}p_{L,h}, p_{L,h})_{0,\Omega} \\
&\leq \max \tau_l^{-2} \|\mathbf{y}_h\|_{M_h}^2,
\end{aligned}$$

the assertion (3.3b) follows with  $C_1 = 3C_{\text{inv}v_{\text{p,ref}}}C_{\text{param}}$  and  $C_2 = \frac{1}{2} \max \tau_l^{-1}$ .  $\square$

In order to achieve well-balanced convergence in space and time, we choose time step and mesh size such that  $v_p\Delta t = \mathcal{O}(h)$ . Moreover, in our application the dissipation rate  $\Delta t(D_h\mathbf{y}_h, \mathbf{y}_h)_{0,\Omega}$  is much smaller than the energy  $(M_h\mathbf{y}_h, \mathbf{y}_h)_{0,\Omega}$  so that  $\Delta t\tau_l^{-1}$  is small. Then, we obtain from (3.3) that the linear system (3.2) with block-diagonal preconditioner  $M_h^{-1}$  is well-conditioned with  $C = \mathcal{O}(1)$ . Hence, the error of GMRES iterates  $\mathbf{y}_h^{n,0}, \mathbf{y}_h^{n,1}, \mathbf{y}_h^{n,2}, \dots$  is bounded by

$$\|\mathbf{y}_h^{n,k} - \mathbf{y}_h^n\|_{M_h} \leq C(1 - C^{-2})^{k/2} \|\mathbf{y}_h^{n,0} - \mathbf{y}_h^n\|_{M_h},$$

see [26, Chap. 6.11.4], i.e., every time step can be computed within a few GMRES steps.

The convergence of the DG method can be estimated by the equivalent space-time scheme, see [8, Sect. 3]: Let  $\mathbf{U}_h \subset H^1(0, T; \mathbf{Y}_h)$  be the discrete space of linear splines in time with respect to the time grid  $0 < t_1 < \cdots < t_N$ . Further,  $\mathbf{y}_h(0) = \mathbf{0}$  for  $\mathbf{y}_h \in \mathbf{U}_h$ . Let  $\mathbf{y}_h^{\text{midpoint}} \in \mathbf{U}_h$  be the discrete solution defined by  $\mathbf{y}_h^n \in \mathbf{Y}_h$  solving (3.2). Provided that the solution  $\mathbf{y}^{\text{sol}} \in \mathbf{U}$  of  $\mathbf{L}\mathbf{y} = \mathbf{b}$  is in  $H^2((0, T); H^{q+1}(\Omega; \mathbb{R}^{d+1+L}))$ ,  $q \leq k$ , we obtain the estimate in space and time

$$\|\mathbf{y}^{\text{sol}} - \mathbf{y}_h^{\text{midpoint}}\|_{L_h} \leq C_1(\Delta t + h^q) \left( \|\partial_t^2 \mathbf{y}^{\text{sol}}\|_{0,(0,T)\times\Omega} + \|D^{q+1} \mathbf{y}^{\text{sol}}\|_{0,(0,T)\times\Omega} \right),$$

compare [7, Thm. 5]. The above error is expressed in the weighted graph norm of the operator  $L_h$ :

$$\|z\|_{L_h} = \sqrt{(M_h z, z)_{0,(0,T)\times\Omega} + (M_h^{-1}L_h z, L_h z)_{0,(0,T)\times\Omega}}, \quad z \in H^1(0, T; \mathbf{Y}_h).$$

*Remark 3.2.* Extrapolation of the midpoint rule

$$(3.4) \quad \mathbf{y}_h^{n,\text{ex}} = \frac{4}{3} \boldsymbol{\eta}_{\Delta t/2} - \frac{1}{3} \boldsymbol{\eta}_{\Delta t},$$

yields a scheme of higher order where  $\boldsymbol{\eta}_{\Delta t}$  and  $\boldsymbol{\eta}_{\Delta t/2}$  are defined by

$$\begin{aligned} \left(\mathbf{M}_h + \frac{\Delta t}{2}\mathbf{A}_h + \frac{\Delta t}{2}\mathbf{D}_h\right)\boldsymbol{\eta}_{\Delta t}^n &= \left(\mathbf{M}_h - \frac{\Delta t}{2}\mathbf{A}_h - \frac{\Delta t}{2}\mathbf{D}_h\right)\mathbf{y}_h^{n-1,\text{ex}} + \Delta t \mathbf{b}_h^{n-1/2}, \\ \left(\mathbf{M}_h + \frac{\Delta t}{4}\mathbf{A}_h + \frac{\Delta t}{4}\mathbf{D}_h\right)\boldsymbol{\eta}_{\Delta t/2}^{n-1/2} &= \left(\mathbf{M}_h - \frac{\Delta t}{4}\mathbf{A}_h - \frac{\Delta t}{4}\mathbf{D}_h\right)\mathbf{y}_h^{n-1,\text{ex}} + \frac{\Delta t}{2} \mathbf{b}_h^{n-1/4}, \\ \left(\mathbf{M}_h + \frac{\Delta t}{4}\mathbf{A}_h + \frac{\Delta t}{4}\mathbf{D}_h\right)\boldsymbol{\eta}_{\Delta t/2}^n &= \left(\mathbf{M}_h - \frac{\Delta t}{4}\mathbf{A}_h - \frac{\Delta t}{4}\mathbf{D}_h\right)\boldsymbol{\eta}_{\Delta t/2}^{n-1/2} + \frac{\Delta t}{2} \mathbf{b}_h^{n-3/4}. \end{aligned}$$

Evaluating the corresponding Butcher tableau (Table 1) we observe that the discrete scheme is of order 4 in time.

1/2	1/2		
1/4	0	1/4	
3/4	0	1/2	1/4
	-1/3	2/3	2/3

Table 1: Butcher tableau of the extrapolated midpoint rule.

**4. Accuracy of the DG approximation.** In two numerical tests we evaluate the convergence properties in space and time of the discrete scheme introduced in section 3. The first test considers the optimal case with known smooth solution, and then a more realistic setting is tested where estimates for the convergence rates are derived by extrapolation.

**4.1. Convergence in homogeneous media.** For the first test we fix  $\omega > 0$  and we choose  $\kappa_0, \tau_p > 0$  such that  $(1 + L\tau_p)\kappa_0 = 1$ ,  $\varrho = 1$ .

In this setting,

$$\mathbf{y}^{\text{sol}}(t, x) = \begin{pmatrix} (t-1)\omega \cos(\omega x_1) \sin(\omega x_2) \\ (t-1)\omega \cos(\omega x_2) \sin(\omega x_1) \\ \kappa_0 \sin(\omega x_1) \sin(\omega x_2) \\ \tau_p \kappa_0 \sin(\omega x_1) \sin(\omega x_2) \\ \dots \\ \tau_p \kappa_0 \sin(\omega x_1) \sin(\omega x_2) \end{pmatrix}$$

solves the visco-acoustic equation for given right-hand side

$$\mathbf{b}(t, x) = \begin{pmatrix} \mathbf{0} \\ 2(t-1)\omega^2 \sin(\omega x_2) \sin(\omega x_1) \\ 2(t-1)\omega^2 \sin(\omega x_1) \sin(\omega x_2) + \sin(\omega x_1) \sin(\omega x_2)/\tau_1 \\ \dots \\ 2(t-1)\omega^2 \sin(\omega x_1) \sin(\omega x_2) + \sin(\omega x_1) \sin(\omega x_2)/\tau_L \end{pmatrix},$$

Dirichlet boundary values  $\mathbf{y}^{\text{sol}}(t)|_{\partial\Omega}$ , and initial values  $\mathbf{y}^{\text{sol}}(0)$ .

For our calculations we use  $L = 1$ ,  $\kappa_0 = 0.85$ ,  $\tau_p = 3/17$  ( $Q \approx 2/\tau_p \approx 11$ ),  $\tau_1 = 0.0001$  and  $\omega = \frac{\pi}{4}$ ,  $\Omega = (0, 4)^2$ , and  $T = 1$ . The initial value  $\mathbf{y}_h^0$  for our calculations is the interpolation of  $\mathbf{y}^{\text{sol}}(0)$  in the finite element space  $\mathbf{Y}_h$ . As time integrator we use the implicit midpoint rule (3.2) with a fixed step size  $\Delta t = 0.001$ . Since the solution is linear in time, the time integrator is exact, and we only study the convergence in space.

We start with a coarse mesh  $\mathcal{K}_0$  with  $h_0 = 2^{-1}$  resulting to  $h_j = 2^{-j}h_0$  on mesh level  $j$  obtained by uniform refinement. Let  $\mathbf{y}_{j,k}$  be the numerical solution on level  $j$



for the DG scheme with polynomial degree  $k$ . We observe optimal convergence  $\mathcal{O}(h_j^{k+1})$ , cf. [Table 2](#).

$j$	$\ \mathbf{y}_{j,0} - \mathbf{y}^{\text{sol}}\ _{0,\mathcal{C}}$	$\ \mathbf{y}_{j,1} - \mathbf{y}^{\text{sol}}\ _{0,\mathcal{C}}$	$\ \mathbf{y}_{j,2} - \mathbf{y}^{\text{sol}}\ _{0,\mathcal{C}}$	$\ \mathbf{y}_{j,3} - \mathbf{y}^{\text{sol}}\ _{0,\mathcal{C}}$
0	$5.539 \cdot 10^{-1}$	$2.807 \cdot 10^{-1}$	$9.324 \cdot 10^{-3}$	$4.886 \cdot 10^{-4}$
1	$3.004 \cdot 10^{-1}$ <b>0.88</b>	$7.288 \cdot 10^{-2}$ <b>1.95</b>	$1.059 \cdot 10^{-3}$ <b>3.14</b>	$3.033 \cdot 10^{-5}$ <b>4.01</b>
2	$1.563 \cdot 10^{-1}$ <b>0.94</b>	$1.837 \cdot 10^{-2}$ <b>1.99</b>	$1.262 \cdot 10^{-4}$ <b>3.07</b>	$1.843 \cdot 10^{-6}$ <b>4.04</b>
3	$7.971 \cdot 10^{-2}$ <b>0.97</b>	$4.595 \cdot 10^{-3}$ <b>2.00</b>	$1.542 \cdot 10^{-5}$ <b>3.03</b>	$1.145 \cdot 10^{-7}$ <b>4.01</b>

Table 2: Convergence in space for different polynomial degrees  $k$ : the rate on level  $j$  is computed as  $\log_2 \frac{\|\mathbf{y}_{j,k} - \mathbf{y}^{\text{sol}}\|_{0,\mathcal{C}}}{\|\mathbf{y}_{j+1,k} - \mathbf{y}^{\text{sol}}\|_{0,\mathcal{C}}}$  and written in bold face in between the involved  $j$ -rows. The norms are evaluated on the space time cylinder  $\mathcal{C} = (0, T) \times \Omega$ .

**4.2. A test configuration for an application in seismics.** Now we study a configuration which is close to our intended real application. This configuration allows us to calibrate the choice of the discretization parameters with respect to the evaluation of seismograms in the inverse algorithm. Thus, our numerical solver is sufficiently accurate in time and in space.

We consider the square  $\Omega = (0, 2000)^2$  (our length unit is meter) and we select piecewise constant material parameters  $\mu = (\varrho, v_p, \tau_p) \in \mathbf{P}_h$ :

$$v_p(x) = \begin{cases} 1.05v_{\text{bg}}, & x \in \Delta_1 = (900, 1100) \times (600, 800), \\ v_{\text{bg}}, & \text{else,} \end{cases}$$

$$\varrho(x) = \begin{cases} 1.05\varrho_{\text{bg}}, & x \in \Delta_2 = (900, 1100) \times (900, 1100), \\ \varrho_{\text{bg}}, & \text{else,} \end{cases}$$

$$\tau_p(x) = \begin{cases} 4.00\tau_{\text{bg}}, & x \in \Delta_3 = (900, 1100) \times (1200, 1400), \\ \tau_{\text{bg}}, & \text{else.} \end{cases}$$

The background values can be found in [Figure 1](#). Further we choose  $L = 5$  damping mechanisms and use the relaxation times obtained by the extended  $\tau$ -method [\[3\]](#) to approximate a constant quality factor  $Q = 15$  in the frequency range 1 Hz–50 Hz. Numerical values are listed in [Figure 1](#). A plot of the realized  $Q$  [\(2.2b\)](#) with these relaxation times can be seen in [Figure 2](#).

The source is modeled by a Ricker wavelet  $a \in C^\infty(\mathbb{R})$  at the position  $x_{\text{source}} \in \Omega$ . This corresponds to the right-hand side  $\mathbf{b}$  given by

$$(4.1) \quad \mathbf{b} = (0, 0, b_p, \mathbf{0}), \quad b_p(t, x) = a(t) \delta_{x_{\text{source}}}(x), \quad t \in (0, T), \quad x \in \Omega,$$

where the Ricker wavelet is chosen with central frequency  $\omega_0/(2\pi) = 25$  Hz and a small shift away from the origin

$$a(t) = \alpha(t - 3\pi/\omega_0) \quad \text{with} \quad \alpha(t) = (1 - \omega_0^2 t^2/2) \exp(-(\omega_0 t/2)^2), \quad t \in \mathbb{R}.$$

The geometric setting of our test scenario can be seen in [Figure 1](#). To model the measurements, we use in this subsection only one source at  $x_{\text{source}} = (600, 1370)$  (indicated by shot<sub>4</sub> in [Figure 1](#)) and  $S = 64$  equidistant measurement points (receivers)

$$r_s = (1350, 370 + 20s) \in \Omega, \quad s = 0, \dots, 63,$$

on the line  $\{1350\} \times [370, 1620]$ , see [Figure 1](#).

The linear measurement operator  $\Psi_{h,s} \in \mathcal{L}(\mathbf{Y}_h, \mathbb{R})$  at receiver  $r_s$  is a point evaluation of the pressure  $p_h = p_{0,h} + \dots + p_{L,h}$ , that is,

$$\Psi_{h,s}(\mathbf{v}_h, p_{0,h}, \dots, p_{L,h}) = p_h(r_s), \quad (\mathbf{v}_h, p_{h,0}, \dots, p_{h,L}) \in \mathbf{Y}_h.$$

*Remark 4.1.* The right hand side  $\mathbf{b}_h$  of the discretized problem (3.1) can be evaluated analytically for the source term from (4.1): for any test function  $\phi_h \in \mathbf{Y}_h$  we have that

$$(\mathbf{b}(t), \phi_h)_{0,\Omega} = \alpha(t) \phi_{h,3}(x_{\text{source}})$$

where  $\phi_{h,3}$  is the third component of  $\phi_h$ .

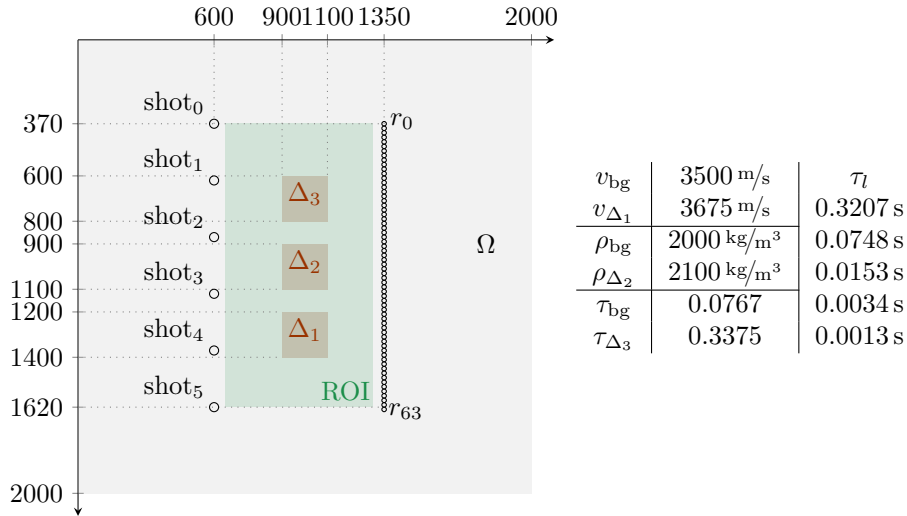


Fig. 1: Geometry, source positions, inclusions, and receiver positions for the convergence test. The region of interest (ROI) for FWI is indicated as light green rectangle.

For the numerical test we use an extended domain  $\Omega^{\text{ext}} \supset \Omega$  such that waves with speed 3500 m/s can never reach any receiver after reflecting at a boundary. In the DG finite element space we fix the polynomial degree  $k = 2$  for all calculations below. Our coarse mesh has mesh size  $h_0 = 50$  and the meshes on the higher levels are obtained by bisecting the mesh size. We emphasize that the initial mesh and hence all finer meshes are aligned with the supports  $\Delta_i$ ,  $i = 1, 2, 3$ , of the three inclusions.

In time we use the uniform meshes with time step sizes  $\Delta t_m = 0.00064 \cdot 2^{-m}$ ,  $m = -1, 0, 1, 2, 3, \dots$ . The linear system in every time step (3.2) is solved approximately by a preconditioned GMRES iteration with a reduction factor<sup>2</sup>  $10^{-4}$ . In our tests 2–5 GMRES steps are required in average.

Let  $\mathbf{y}_{j,m} \in \mathbf{U}_h$  be the DG solution with level  $j$  in space and level  $m$  in time. Then the numerically computed seismogram  $\mathbf{s}_{j,m} \in \mathbf{H}^1(0, T; \mathbb{R}^{64})$  is given by

$$(4.2) \quad (\mathbf{s}_{j,m})_s = \Psi_{h,s} \mathbf{y}_{j,m}, \quad s = 0, \dots, 63,$$

<sup>2</sup>The iteration is stopped when the relative residuals of the GMRES iterates are below the reduction factor.

$j$	$f_{j,0}$	$f_{j,1}$	$f_{j,2}$	$f_{j,3}$
0	$2.106 \cdot 10^{-1}$ <b>3.34</b>	$2.097 \cdot 10^{-1}$ <b>3.35</b>	$2.095 \cdot 10^{-1}$ <b>3.35</b>	$2.094 \cdot 10^{-1}$ <b>3.35</b>
1	$2.080 \cdot 10^{-2}$ <b>3.04</b>	$2.062 \cdot 10^{-2}$ <b>3.01</b>	$2.058 \cdot 10^{-2}$ <b>3.02</b>	$2.058 \cdot 10^{-2}$ <b>3.03</b>
2	$2.534 \cdot 10^{-3}$	$2.559 \cdot 10^{-3}$	$2.534 \cdot 10^{-3}$	$2.534 \cdot 10^{-3}$

Table 3: Convergence of seismograms in space for fixed polynomial degree  $k = 2$ , where  $f_{j,m} = \frac{\|\mathbf{s}_{j+1,m} - \mathbf{s}_{j,m}\|_{0,(0,T)}}{\|\mathbf{s}_{3,3}\|_{0,(0,T)}}$  are normalized differences of calculated seismograms of two consecutive space levels. The rate on level  $j$  in column  $m$  is computed as  $\log_2 \frac{f_{j,m}}{f_{j+1,m}}$  and written in bold face between the involved  $j$ -rows. The implicit midpoint rule was used as time-stepping scheme.

where the discrete measurement operator acts only on the spatial variable of  $\mathbf{y}_{j,m}$ . Hence,  $(\mathbf{s}_{j,m})_s$  is a linear spline function in time.

We observe cubic convergence in space (Table 3) and using the implicit midpoint rule we even obtain quadratic convergence in time (Table 4).

To estimate the approximation error in the seismograms we use extrapolation to the limit<sup>3</sup>, see, e.g., [10, Chap. 4.2.8]: Let  $\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2$  be three seismograms computed on successive discretization levels. From  $\mathbf{s}_2 - \mathbf{s}_1$  and  $\mathbf{s}_1 - \mathbf{s}_0$  the numerically observed error reduction rate  $r$  is computed, which then defines the extrapolated seismogram by  $\mathbf{s}^{\text{ex}} = \frac{r}{r-1} \mathbf{s}_2 - \frac{1}{r-1} \mathbf{s}_1$  used as a substitute for the unknown exact seismogram.

We extrapolate first in space. Since we observe cubic convergence in space (Table 3) we obtain with  $r = 2^3$

$$\mathbf{s}_m^{\text{ex}} = \frac{8}{7} \mathbf{s}_{3,m} - \frac{1}{7} \mathbf{s}_{2,m} \quad m = 0, 1, 2, 3,$$

for different time levels. Now we interpolate in time where we observe quadratic order (Table 4), and with  $r = 2^2$  we obtain

$$\mathbf{s}^{\text{ex}} = \frac{4}{3} \mathbf{s}_3^{\text{ex}} - \frac{1}{3} \mathbf{s}_2^{\text{ex}}.$$

With the extrapolated seismogram  $\mathbf{s}^{\text{ex}}$  we estimate that the normalized error for the finest seismogram  $\mathbf{s}_{3,3}$  in space and in time is smaller than 0.05%, see Table 5. In the following section we use the space discretization at level  $j = 1$  and the time step  $\Delta t = 0.0005 \in (\Delta t_0, \Delta t_1)$ . Interpolating the corresponding values of Table 5 we expect a relative error of roughly 2% for this space and time discretization.

We also computed the seismograms with the extrapolated implicit midpoint rule, see Table 6. If we lower the reduction factor to  $10^{-8}$  in GMRES we clearly observe fourth-order convergence which we do not observe with the reduction factor  $10^{-4}$ . We only ran one test for this time stepping scheme, since the feasibility of higher order implicit methods in time in the context of FWI is not in the focus here.

<sup>3</sup>The extrapolation technique relies on an error expansion and assumes that the numerical results are in the asymptotic regime which is not guaranteed for our application.

$m$	$g_{0,m}$	$g_{1,m}$	$g_{2,m}$	$g_{3,m}$
0	$2.273 \cdot 10^{-2}$ <b>2.00</b>	$3.164 \cdot 10^{-2}$ <b>2.00</b>	$3.218 \cdot 10^{-2}$ <b>2.01</b>	$3.211 \cdot 10^{-2}$ <b>2.00</b>
1	$5.677 \cdot 10^{-3}$ <b>1.99</b>	$7.900 \cdot 10^{-3}$ <b>2.00</b>	$8.014 \cdot 10^{-3}$ <b>2.00</b>	$8.028 \cdot 10^{-3}$ <b>2.01</b>
2	$1.430 \cdot 10^{-3}$	$1.972 \cdot 10^{-3}$	$1.997 \cdot 10^{-3}$	$1.996 \cdot 10^{-3}$

Table 4: Convergence of seismograms in time for the implicit midpoint rule and fixed polynomial degree  $k = 2$ , where  $g_{j,m} = \frac{\|\mathbf{s}_{j,m+1} - \mathbf{s}_{j,m}\|_{0,(0,T)}}{\|\mathbf{s}_{3,3}\|_{0,(0,T)}}$  are normalized differences of calculated seismograms of two consecutive time levels. The rate on level  $m$  and for fixed  $j$  is computed as  $\log_2 \frac{g_{j,m}}{g_{j,m+1}}$  and written in bold face between the involved  $m$ -rows.

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$\frac{\ \mathbf{s}_m^{\text{ex}} - \mathbf{s}^{\text{ex}}\ _{0,(0,T)}}{\ \mathbf{s}^{\text{ex}}\ _{0,(0,T)}}$
$m = 0$	0.20466	0.02798	0.04056	0.04250	0.04278
$m = 1$	0.21682	0.01609	0.00857	0.01041	0.01069
$m = 2$	0.22018	0.02106	0.00182	0.00239	0.00266
$m = 3$	0.22104	0.02256	0.00241	0.00044	0.00067

Table 5: Normalized error estimates  $\frac{\|\mathbf{s}_{j,m} - \mathbf{s}^{\text{ex}}\|_{0,(0,T)}}{\|\mathbf{s}^{\text{ex}}\|_{0,(0,T)}}$  for different space  $j$  and time levels  $m$ .

$m$	$g_{0,m}$	$g_{1,m}$	$g_{2,m}$	$g_{3,m}$
-1	$2.691 \cdot 10^{-4}$ <b>3.98</b>	$5.087 \cdot 10^{-4}$ <b>3.98</b>	$5.352 \cdot 10^{-4}$ <b>3.97</b>	-
0	$1.703 \cdot 10^{-5}$ <b>3.99</b>	$3.234 \cdot 10^{-5}$ <b>3.99</b>	$3.405 \cdot 10^{-5}$ <b>3.99</b>	$3.402 \cdot 10^{-5}$ <b>3.99</b>
1	$1.069 \cdot 10^{-6}$ <b>4.00</b>	$2.031 \cdot 10^{-6}$ <b>4.00</b>	$2.139 \cdot 10^{-6}$ <b>4.00</b>	$2.137 \cdot 10^{-6}$ <b>4.00</b>
2	$6.674 \cdot 10^{-8}$	$1.273 \cdot 10^{-7}$	$1.339 \cdot 10^{-7}$	$1.337 \cdot 10^{-7}$

Table 6: Convergence of seismograms in time for the extrapolated implicit midpoint rule and fixed polynomial degree  $k = 2$  where  $g_{j,m} = \frac{\|\mathbf{s}_{j,m+1} - \mathbf{s}_{j,m}\|_{0,(0,T)}}{\|\mathbf{s}_{3,3}\|_{0,(0,T)}}$  are normalized differences of calculated seismograms of two consecutive time levels. The rate on level  $m$  and for fixed  $j$  is computed as  $\log_2 \frac{g_{j,m}}{g_{j,m+1}}$  and written in bold face between the involved  $m$ -rows.

**5. Reconstruction.** In this section we present an application of the combination of the inexact Newton-CG method with the DG method as forward solver. The goal is to give a proof of concept. Therefore we keep the configuration, the model and the inversion strategy as elementary as possible to explore the performance of the new combination in visco-acoustic media. We choose the same crosswell scenario as in our test of convergence using the geometry in Figure 1. We assume that the source time function is known and do not apply a source time function inversion. The background model is used as initial model. A multi-scale strategy by sequential frequency filtering is not required as the model perturbations are local and no cycle skipping is observed in the considered frequency range. We also do not apply a preconditioning of the gradient to reduce artifacts around sources and receivers.

In realistic applications there are always more than one source, say, we have

**Algorithm 5.1** KCG-REGINN

---

**Input:**  $\mu^0 \in \mathcal{P}_h$  % starting guess;  $\mathbf{s}_{\text{obs}} \in L_2(0, T; \mathbb{R}^S)^\Sigma$  % seismograms  
**Output:**  $\mu^k \in \mathcal{P}_h$  % approximate solution of (2.14)

- 1:  $k \leftarrow 0$ ,  $\text{shot} \leftarrow 0$ ,
- 2: **while not** termination **do**
- 3:  $\mu_0^k \leftarrow \mu^k$
- 4: **repeat**
- 5:  $\mathbf{r}_{\text{shot}}^k = \mathbf{s}_{\text{obs,shot}} - \Phi_{\text{shot}}(\mu_{\text{shot}}^k)$
- 6: determine  $\vartheta_{\text{shot}}^k$  and  $\text{shot}_{\text{max,shot}}^k$  % according to (2.18) and (2.19)
- 7:  $(\Delta\mu_{\text{shot}}^k, j_k) \leftarrow \text{CG}(\mathbf{r}_{\text{shot}}^k, \mu_{\text{shot}}^k, \vartheta_{\text{shot}}^k, \text{shot}_{\text{max,shot}}^k)$  % call of Algorithm 2.2
- 8:  $\mu_{\text{shot}+1}^k \leftarrow \mu_{\text{shot}}^k + \Delta\mu_{\text{shot}}^k$
- 9:  $\text{shot} \leftarrow \text{shot} + 1$
- 10: **until**  $\text{shot} \geq \Sigma$
- 11:  $\text{shot} \leftarrow 0$
- 12:  $\mu^{k+1} \leftarrow \mu_{n_{\text{shots}}-1}^k$
- 13:  $k \leftarrow k + 1$
- 14: **end while**
- 15: **return**  $\mu^k$

---

$\Sigma \in \mathbb{N}$  sources (we use  $\Sigma = 6$ , see Figure 1). If a source fires we call it a *shot*. For each of our shots we have a forward operator  $\Phi_{\text{shot}}$ ,  $\text{shot} = 0, \dots, \Sigma - 1$ , which maps the model parameters  $\mu$  to the corresponding seismogram  $\mathbf{s}_{\text{shot}}$  (as in (4.2) for each shot). Thus we are lead to  $\Sigma$  subproblems

$$(5.1) \quad \Phi_{\text{shot}}(\mu) = \mathbf{s}_{\text{obs,shot}}, \quad \text{shot} = 0, \dots, \Sigma - 1,$$

where  $\mathbf{s}_{\text{obs,shot}}$  is the observed/measured seismogram.

To reduce the numerical effort per iteration step we provide a Kaczmarz variant of CG-REGINN (Algorithm 2.1) where all equations in (5.1) are processed individually and cyclically by CG-REGINN (Algorithm 5.1). Kaczmarz solvers are well established and explored in the inverse problem community, see, e.g., [18, 12, 11, 21, 22].

In all our experiments KCG-REGINN is stopped after 10 iterations (passes through the while-loop). As  $\Sigma = 6$  the model parameter is updated 60 times.

We work with the same geometrical layout as before, see Figure 1. We also use the same attenuation described there,  $\tau_{\text{bg}} = 0.0767$  and the  $\tau_l$ 's given in Figure 1. Within this setting we approximate a nearly constant  $Q = 15$  in the relevant frequency range given by the center frequency  $\omega_0/(2\pi) = 25$  Hz of the source. The background values  $\rho_{\text{bg}} = 2000$  kg/m<sup>3</sup> and  $v_{\text{bg}} = 3500$  m/s are also those from subsection 4.2.

For the inversion of  $v_p$  we increased the corresponding background value by 5% in  $\Delta_1$ , by 10% in  $\Delta_2$ , and decreased it by 5% in  $\Delta_3$ . For the inversion of  $\tau_p$  the parameter  $\tau_p$  is given by  $\lambda\tau_{\text{bg}}$  for a factor  $\lambda > 0$  such that the approximate values for  $Q$  (2.2b) are (the values for  $\lambda$  are in parentheses): 5(4.4), 10(1.7) and 30(0.45) in  $\Delta_1$ ,  $\Delta_2$ , and  $\Delta_3$ , respectively. The behavior of the  $Q$ 's over frequency can be seen in Figure 2. The numerical values of all involved parameters of our reconstruction experiments are listed in Table 7. The larger contrast in  $\tau_p$  compared to  $v_p$  is needed so that the wave is noticeably attenuated on its short (relative to its wavelength) way from the sources to the receivers.

We apply a monoparameter strategy for each parameter individually, this means only one parameter is reconstructed in every experiment. The starting guess  $\mu^0$  for Algorithm 5.1 was always chosen as the background value of the parameter under consideration.

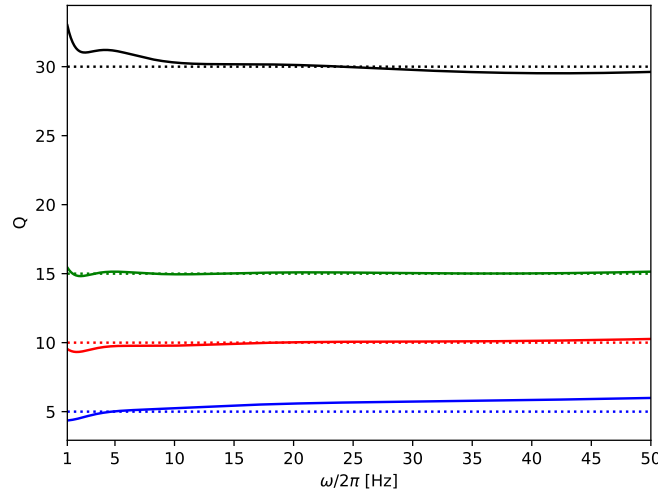


Fig. 2: Approximation of  $Q$  (2.2b) over the frequency range 1 – 50 Hz for the four different material areas in the test configuration, see Figure 1 and right table in Table 7. The dashed lines indicate the constant target values for  $Q$ .

$v_{bg}$	3500 m/s	$\tau_l$	$\tau_{bg}$	0.0767	$\tau_l$
$v_{\Delta_1}$	3675 m/s	0.3207 s	$\tau_{\Delta_1}$	0.3375	0.3207 s
$v_{\Delta_2}$	3850 m/s	0.0748 s	$\tau_{\Delta_2}$	0.0345	0.0748 s
$v_{\Delta_3}$	3325 m/s	0.0153 s	$\tau_{\Delta_3}$	0.1304	0.0153 s
$\rho_{bg}$	2000 kg/m <sup>3</sup>	0.0034 s	$\rho_{bg}$	2000 kg/m <sup>3</sup>	0.0034 s
$\tau_{bg}$	0.0767	0.0013 s	$v_{bg}$	3500 m/s	0.0013 s

Table 7: parameters for the monoparameter reconstructions. Left: parameters for reconstructing  $v_p$ . Right: parameters for reconstructing  $\tau_p$ .

For each parameter setup we invert three different input data sets so that we are able to compare reconstructions with and without inverse crime. Inverse crime (IC) means generating synthetic measurements with the same forward solver used later during inversion. Committing an IC results in ‘over-optimistic’ reconstructions which lack typical artifacts originating from the continuous model, see [5]. For instance, source modeling is a delicate and sensitive task which leads to an additional modeling error causing artifacts which are virtually not visible in reconstructions from IC data. This effect can be seen clearly below in the experiments with respect to the scaling parameter  $\tau_p$ . Moreover, our experiments (Figure 5 and Figure 6 below) demonstrate that IC data disguise the true ill-posed nature of this inverse problem indeed: the results from the strong IC data are better than those by weak or no IC data. Stronger artifacts superimpose the otherwise meaningful outcome.

Our first two data sets are generated by the DG solver introduced in the previous section. The first set was produced using the ansatz space from the inverse solver (strong IC), for the second set we discretized the visco-acoustic equation on a finer grid in space with higher degree polynomials ( $j = 2$ ,  $k = 3$  in the notation of section 4) (weak IC). The third set was obtained by the FDTD code SOFI [4] (no IC). Due to different implementations of the source it was necessary to rescale the FDTD data. For this the whole FDTD dataset was multiplied by the quotient of the maximal amplitude of the traces of the DG and the FDTD data.

The quality of the reconstructions is measured by the relative  $L_2$ - and  $L_\infty$  distances to the analytic ground truth<sup>4</sup>, however, restricted to the region of interest  $\text{ROI} = [650, 1300] \times [370, 1620]$  which is located between sources and receivers (light green rectangle in Figure 1).

**5.1. Inversion for  $v_p$ .** Variations in  $v_p$  produce differences in the phase or traveltime between the predicted and observed waveforms: high and low velocity anomalies lead to an advance and delay of traveltime, respectively (see Figure 4, first row, left). The reconstructions of the P-wave velocity  $v_p$  after 10 iterations are shown in Figure 5. We see that the case of strong IC produces slightly weaker artifacts than the weak IC case. For the case of no IC artifacts are further strengthened. For all cases the reconstructions of  $v_p$  are of good quality. The parameter  $v_p$  is thus not much affected by systematic errors in the observed seismograms.

The seismograms obtained for the final and true models at the receivers  $r_0$ ,  $r_{31}$ , and  $r_{63}$  for the case of no IC are compared in Figure 4. Already after 10 iterations we obtain a nearly perfect fit of seismograms (Figure 4, first row, right) indicating the fast convergence of KCG-REGINN in our mono-parameter crosswell setting. The convergence history is presented in Figure 3 (left):  $L_2$ - and  $L_\infty$ -errors relative to the initial error are plotted as function of the iteration number for the strong and the no IC scenario. We observe a fast and monotone decrease of the  $L_2$ -error in agreement with the theoretic results [19, Thm 3.1]. Only 10 iterations were required to achieve a stable result.

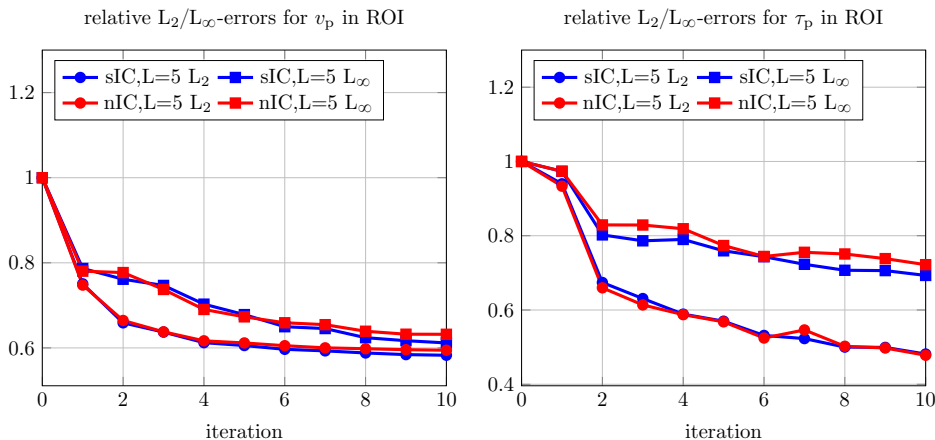


Fig. 3: Relative  $L_2/L_\infty$  errors in  $\text{ROI} = [650, 1300] \times [370, 1620]$  for the strong IC (sIC) and no IC (nIC) data for  $v_p$  (left) and  $\tau_p$  (right).

**5.2. Inversion for  $\tau_p$ .** The reconstruction of the level of attenuation  $\tau_p$  is generally challenging in practical applications due to the presence of coherent noise which mainly affects the recorded amplitudes (and not the phase). Those amplitude fluctuations can easily cause wrong updates of attenuation because attenuation is most sensitive to amplitude variations at different receivers. The effect of attenuation on recorded seismograms can be nicely identified in Figure 4 (second row, left). Here the signal amplitudes of the observed FDTD data increase and decrease when waves travel through a zone with lower and higher attenuation, respectively.

Another challenge associated with the inversion of attenuation is the broad range of values for the level of attenuation  $\tau_p$  that must be expected in real applica-

<sup>4</sup>As our parameters are represented by piece-wise constant functions, see subsection 2.5, we have an analytic formula for the  $L_2$ -norm.

tion. The relative changes in  $\tau_p$  are generally much higher than those in  $v_p$  for the same earth model. To overcome this challenge we reduce the parameter range of  $\tau_p$  and force positivity by the ansatz  $\tau_p = \exp \tilde{\tau}_p$ , see [Remark 2.4](#). The derivatives with respect to this new parameter can be calculated via the chain rule, see [Appendix A](#) for details.

In [Figure 4](#) we compare seismograms for the initial and final model at the receivers  $r_0$ ,  $r_{31}$ , and  $r_{63}$  for the case of no IC. Already after 10 iterations we observe again a nearly perfect data fit ([Figure 4](#), second row, right). This means that the final attenuation model can perfectly explain the observations. The final reconstructions are plotted in [Figure 6](#) for the three cases of IC. In this mono-parameter configuration both positive and negative deviations of the attenuation parameter  $\tau_p$  can be recovered surprisingly well by KCG-REGINN. A good reconstruction is achieved if the deviation from the background is sufficiently large, i.e. if the effect in the observed seismograms exceed the systematic errors produced by the forward solver. Compared to the reconstructions of  $v_p$  ([Figure 5](#)), the reconstructions of  $\tau_p$  are affected more severely by IC, i.e. systematic amplitude errors in the used forward solvers. The strongest artifacts develop in the vicinity of the source locations when the FDTD data (no IC) is inverted. These strong local artifacts are visible in the differences of the two model norms shown in [Figure 3](#) right. At higher iteration numbers the  $L_\infty$  error becomes noticeably larger than the  $L_2$  error. This is not the case in the reconstructions of  $v_p$  (from about the sixth iteration) due to the absence of strong local artifacts.

The development of  $\tau_p$  for all iterations of KCG-REGINN for the case of no IC data can be studied in [Figure 7](#). We see the nice and smooth convergence of the reconstructed model to the true model in the center between the two boreholes. Unfortunately, the artifacts around the sources develop simultaneously. The artifacts around the sources also appear in real data applications and can be reduced by a gradient taper and a simultaneous source time function inversion.

**6. Conclusions.** In this work we presented a modified formulation of the system of forward equations for visco-acoustic media which allows the derivation of the corresponding adjoint equations in an elegant way. The new equations are solved by a discontinuous Galerkin method which shows sufficient accuracy in space and time for a simple but realistic geophysical application, a cross-well scenario. In this cross-well setting the inexact Newton-CG methods converges fast to the true models of attenuation and velocity within a few iterations only. Therefore, the results proof the concept of the suggested framework to be flexible and efficient in realistic applications. Our next goal is to develop this strategy towards more realistic applications. For this purpose we will first extend the framework to visco-elastic media to account also for shear waves and surface waves which generally appear in land seismic recordings. In the long term we plan to apply the new framework to land seismic field data where the data is affected by topographic variations of the earth surface. In this context the inherent advantages of the DG method in simulation strong contrast interface such as free surface topography become most beneficial.

**Appendix A. Formulas for the adjoint Fréchet operator under forced positivity.** If we consider the ansatz  $\rho = \exp \tilde{\rho}$ ,  $\tau_p = \exp \tilde{\tau}_p$ ,  $v_p = \exp \tilde{v}_p$  and invert for  $\tilde{\rho}$ ,  $\tilde{\tau}_p$ ,  $\tilde{v}_p$  we make sure that only positive values are reconstructed. This ansatz gives the parametrization

$$\begin{aligned} & (\rho, \kappa_0^{-1}, \kappa_1^{-1}, \dots, \kappa_L^{-1}) \\ &= \left( \exp(\tilde{\rho}), \frac{1 + \alpha \exp(\tilde{\tau}_p)}{\exp(2\tilde{v}_p + \tilde{\rho})}, \frac{1 + \alpha \exp(\tilde{\tau}_p)}{\exp(2\tilde{v}_p + \tilde{\rho} + \tilde{\tau}_p)}, \dots, \frac{1 + \alpha \exp(\tilde{\tau}_p)}{\exp(2\tilde{v}_p + \tilde{\rho} + \tilde{\tau}_p)} \right). \end{aligned}$$



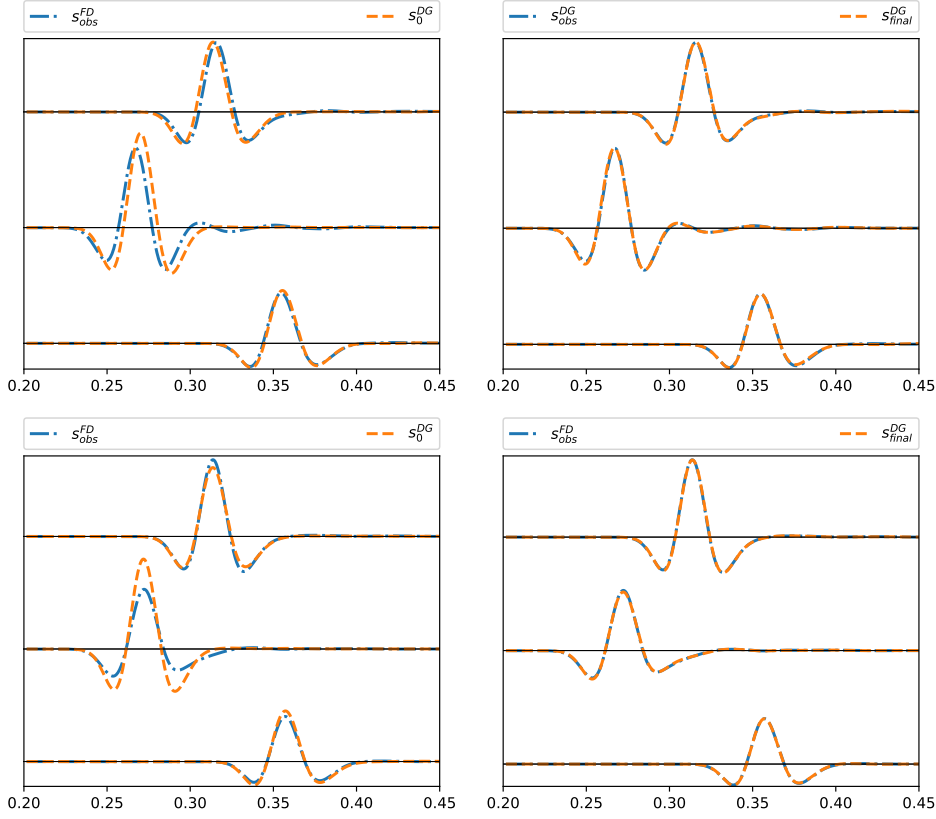


Fig. 4: Comparison of the “no inverse crime” measured (blue) and reconstructed (orange) seismograms from shot 3 for receivers  $r_{63}$ ,  $r_{31}$ , and  $r_0$  (top to bottom). First row: reconstruction experiment for  $v_p$ ; second row: reconstruction experiment for  $\tau_p$ . Left: initial seismograms (orange); right: final reconstructed seismograms (orange).

If we follow the steps of [section 2](#) we get the following formulas for the adjoint operator, compare [\(2.21\)](#),

$$\begin{aligned} \tilde{\rho}_K^{\text{ad}} &= \frac{1}{|K|} \int_0^T \int_K \left( -\exp(\tilde{\rho}) \partial_t \mathbf{v}^{\text{sol}} \cdot \mathbf{w}^{\text{ad}} + \frac{1 + \alpha \exp(\tilde{\tau}_p)}{\exp(2\tilde{v}_p + \tilde{\rho} + \tilde{\tau}_p)} \left( \exp(\tilde{\tau}_p) p_0^{\text{sol}} q_0^{\text{ad}} \right. \right. \\ &\quad \left. \left. + \frac{1}{\tau_{p,K}} \sum_{l=1}^L (\partial_t p_l^{\text{sol}} + \frac{1}{\tau_l} p_l^{\text{sol}}) q_l^{\text{ad}} \right) \right) dx dt, \\ \tilde{v}_{p,K}^{\text{ad}} &= \frac{1}{|K|} \int_0^T \int_K \left( \frac{2 + 2\alpha \exp(\tilde{\tau}_p)}{\exp(2\tilde{v}_p + \tilde{\rho} + \tilde{\tau}_p)} \left( \exp(\tilde{\tau}_p) p_0^{\text{sol}} q_0^{\text{ad}} \right. \right. \\ &\quad \left. \left. + \frac{1}{\tau_{p,K}} \sum_{l=1}^L (\partial_t p_l^{\text{sol}} + \frac{1}{\tau_l} p_l^{\text{sol}}) q_l^{\text{ad}} \right) \right) dx dt, \\ \tilde{\tau}_{p,K}^{\text{ad}} &= \frac{1}{|K|} \int_0^T \int_K \left( -\frac{\alpha \exp(\tilde{\tau}_p)}{\exp(2\tilde{v}_p + \tilde{\rho})} p_0^{\text{sol}} q_0^{\text{ad}} \right. \end{aligned}$$

$$+ \frac{1}{\exp(2\tilde{v}_p + \tilde{\rho} + \tilde{\tau}_p)} \left( \sum_{l=1}^L (\partial_t p_l^{\text{sol}} + \frac{1}{\tau_l} p_l^{\text{sol}}) q_l^{\text{ad}} \right) dx dt .$$

At the end of the inversion we can insert the reconstructed parameters into the exponential function and retrieve  $\rho$ ,  $v_p$ ,  $\tau_p$ .

There is an additional benefit from this re-parametrization: it rescales the original parameters via the logarithm and therefore reduces the absolute and relative difference between the parameters, for instance,  $3500/0.02$  is of order 6 whereas  $\log 3500/\log 0.02$  is of order 1. Huge differences in scale are typically encountered in multiparameter inversion [15].

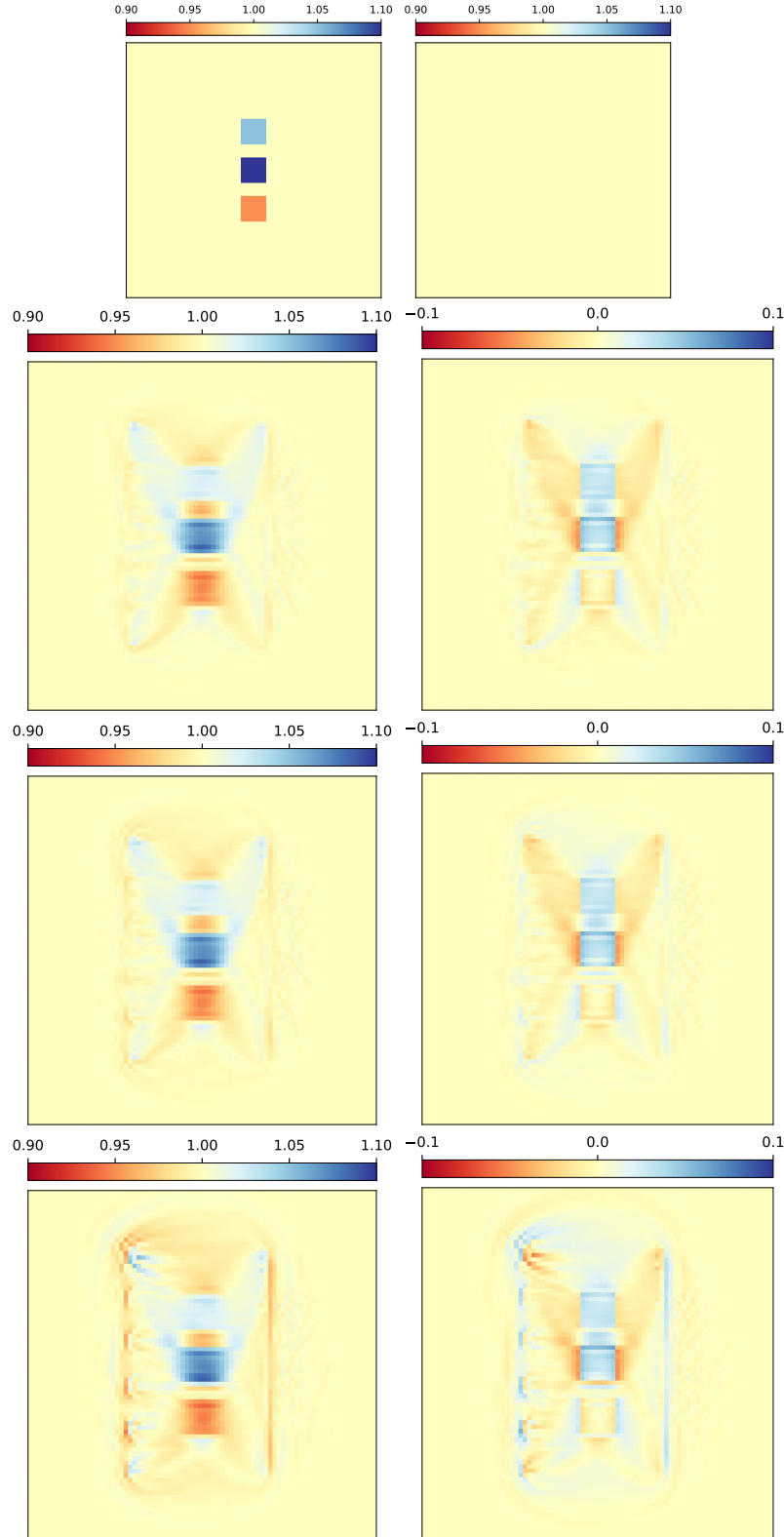


Fig. 5: Inversion for  $v_p$ . First row: ground truth  $v_p^{\text{true}}$  (left) and initial value  $v_p^0$  (right). Second to fourth row: results for strong IC data, weak IC data, and FDTD data. Left: 10-th iterate  $v_p^{10}$  of KCG-REGINN. Right: error  $v_p^{10} - v_p^{\text{true}}$ .

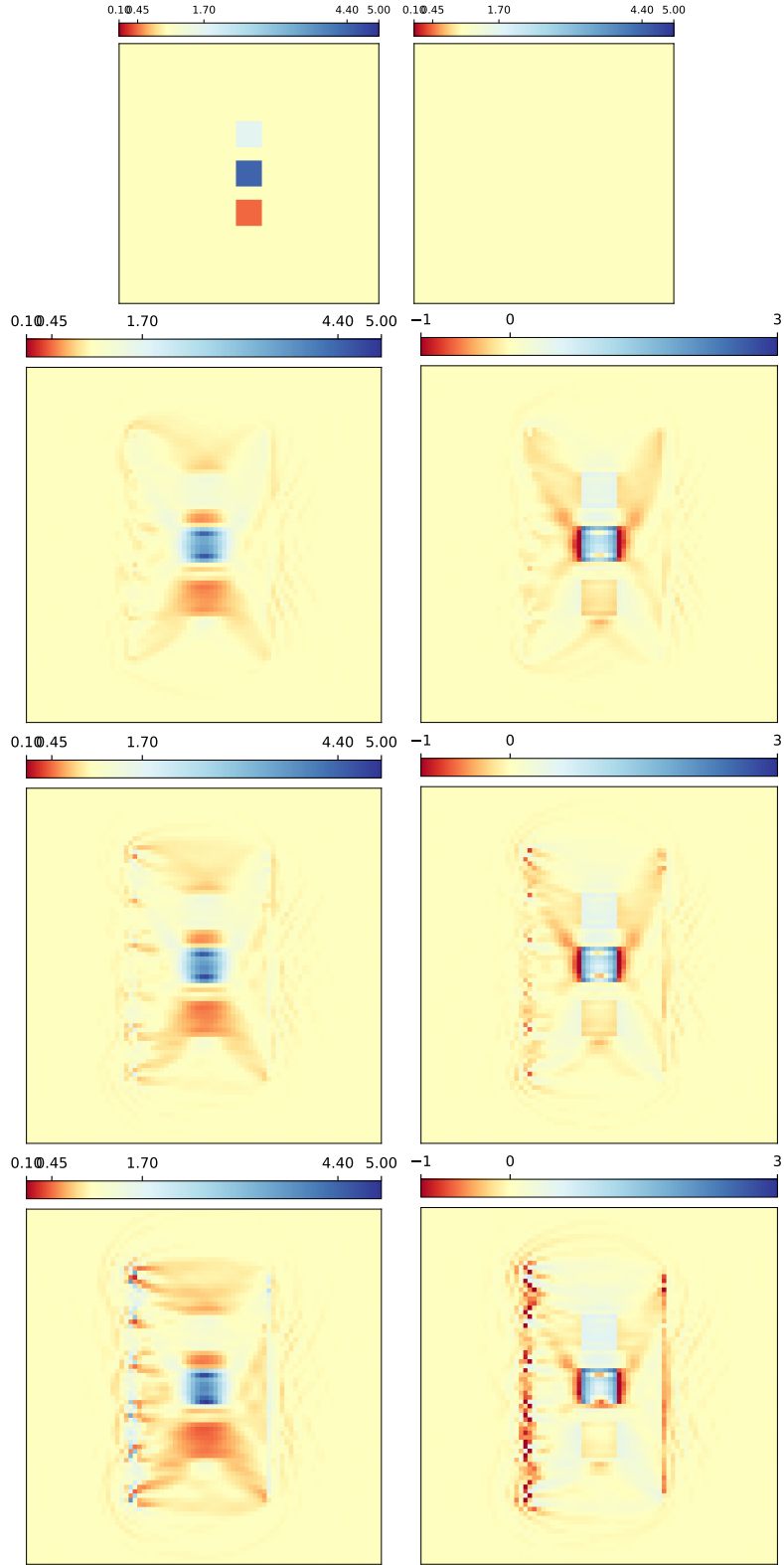


Fig. 6: Inversion for  $\tau_p$ . First row: ground truth  $\tau_p^{\text{true}}$  (left) and initial value  $\tau_p^0$  (right). Second, third and fourth row: results for strong IC data, weak IC data and no IC data. Left: 10-th iterate  $\tau_p^{10}$  of KCG-REGINN. Right: error  $\tau_p^{10} - \tau_p^{\text{true}}$ . In this experiment the Background is  $Q = 15$  and  $L = 5$ .

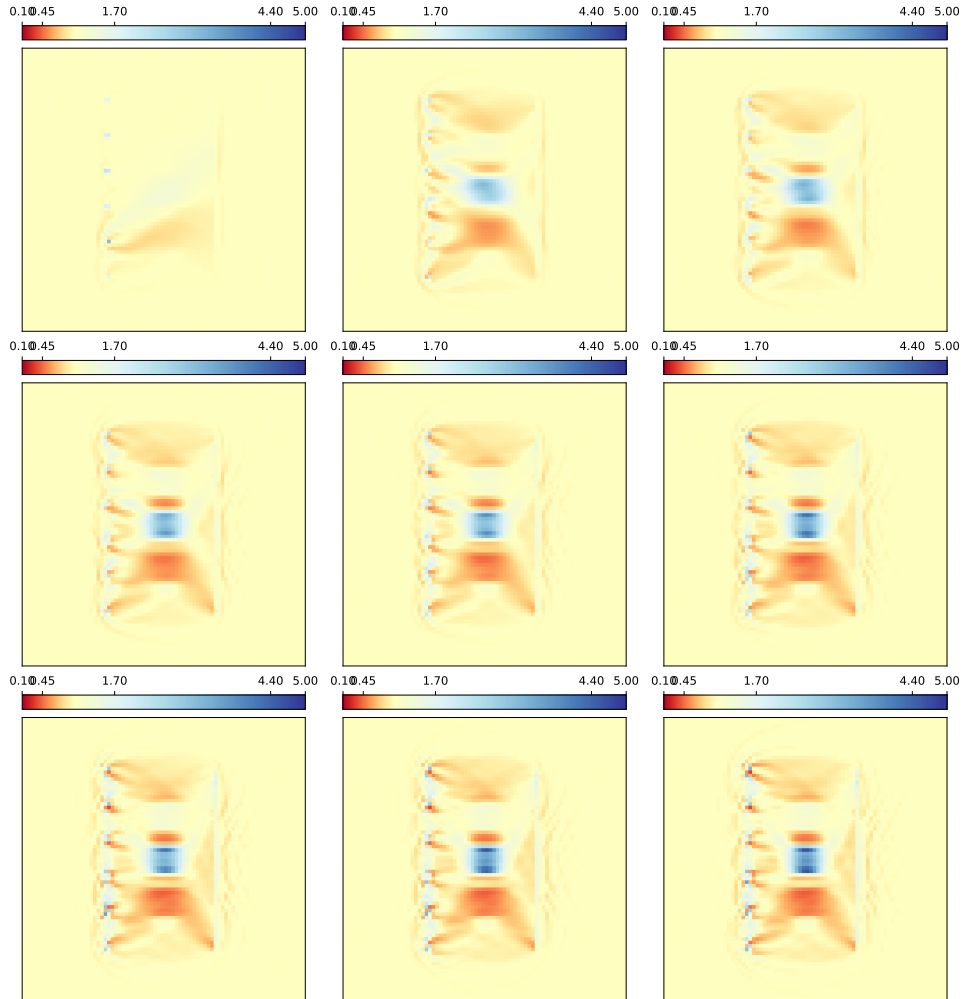


Fig. 7: First 9 iterates of KCG-REGINN for  $\tau_p$  with no IC data. For the 10-th iterate see [Figure 6](#) (4-th row, left).

## REFERENCES

- [1] T. BAI, I. TSVANKIN, AND X. WU, *Waveform inversion for attenuation estimation in anisotropic media*, GEOPHYSICS, 82 (2017), pp. WA83–WA93, <https://doi.org/10.1190/geo2016-0596.1>.
- [2] J. O. BLANCH, J. O. A. ROBERTSSON, AND W. W. SYMES, *Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique*, GEOPHYSICS, 60 (1995), pp. 176–184, <https://doi.org/10.1190/1.1443744>.
- [3] T. BOHLEN, *Interpretation of Measured Seismograms by Means of Viscoelastic Finite Difference Modelling*, PhD thesis, Kiel University, 1998, <https://bit.ly/2LMOSWr>. In German.
- [4] T. BOHLEN, *Parallel 3-D viscoelastic finite difference seismic modelling*, Computers and Geosciences, 28 (2002), pp. 887 – 899, <http://www.sciencedirect.com/science/article/pii/S0098300402000067>.
- [5] D. COLTON AND R. KRESS, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer, 2013, <https://doi.org/10.1007/978-1-4614-4942-3>.
- [6] D. A. DI PIETRO AND A. ERN, *Mathematical aspects of discontinuous Galerkin methods*, vol. 69, Springer Science & Business Media, 2011.
- [7] W. DÖRFLER, S. FINDEISEN, AND C. WIENERS, *Space-time discontinuous Galerkin discretizations for linear first-order hyperbolic evolution systems*, Comput. Methods Appl. Math., 16 (2016), pp. 409–428, <https://doi.org/10.1515/cmam-2016-0015>.
- [8] W. DÖRFLER, S. FINDEISEN, C. WIENERS, AND D. ZIEGLER, *Parallel adaptive discontinuous Galerkin discretizations in space and time for linear elastic and acoustic waves*, in Space-Time Methods. Applications to Partial Differential Equations, U. Langer and O. Steinbach, eds., vol. 21 of Radon Series on Computational and Applied Mathematics, Walter de Gruyter, 2018.
- [9] G. FABIEN-OUELLET, E. GLOAGUEN, AND B. GIROUX, *Time domain viscoelastic full waveform inversion*, Geophysical Journal International, 209 (2017), pp. 1718–1734, <https://dx.doi.org/10.1093/gji/ggx110>.
- [10] W. GANDER, M. J. GANDER, AND F. KWOK, *Scientific computing*, vol. 11 of Texts in Computational Science and Engineering, Springer, Cham, 2014, <https://doi.org/10.1007/978-3-319-04325-8>.
- [11] M. HALTMEIER, R. KOWAR, A. LEITÃO, AND O. SCHERZER, *Kaczmarz methods for regularizing nonlinear ill-posed equations. II. Applications*, Inverse Probl. Imaging, 1 (2007), pp. 507–523, <https://doi.org/10.3934/ipi.2007.1.507>.
- [12] M. HALTMEIER, A. LEITÃO, AND O. SCHERZER, *Kaczmarz methods for regularizing nonlinear ill-posed equations. I. Convergence analysis*, Inverse Probl. Imaging, 1 (2007), pp. 289–298, <https://doi.org/10.3934/ipi.2007.1.289>.
- [13] J. S. HESTHAVEN AND T. WARBURTON, *Nodal discontinuous Galerkin Methods*, Springer, 2008.
- [14] M. HOCHBRUCK, T. PAZUR, A. SCHULZ, E. THAWINAN, AND C. WIENERS, *Efficient time integration for discontinuous Galerkin approximations of linear wave equations*, ZAMM, 95 (2015), pp. 237–259, <https://doi.org/10.1002/zamm.201300306>.
- [15] R. KAMEI AND R. G. PRATT, *Inversion strategies for visco-acoustic waveform inversion*, Geophysical Journal International, 194 (2013), pp. 859–884, <https://doi.org/10.1093/gji/ggt109>.
- [16] A. KIRSCH AND A. RIEDER, *Inverse problems for abstract evolution equations with applications in electrodynamics and elasticity*, Inverse Problems, 32 (2016), pp. 085001, 24, <https://doi.org/10.1088/0266-5611/32/8/085001>.
- [17] A. KIRSCH AND A. RIEDER, *Inverse problems for abstract evolution equations II: higher order differentiability for viscoelasticity*, SIAM J. Appl. Math., 79 (2019), pp. 2639–2662, <https://doi.org/10.1137/19M1269403>.
- [18] R. KOWAR AND O. SCHERZER, *Convergence analysis of a Landweber-Kaczmarz method for solving nonlinear ill-posed problems*, in Ill-posed and inverse problems, VSP, Zeist, 2002, pp. 253–270.
- [19] A. LECHLEITER AND A. RIEDER, *Towards a general convergence theory for inexact Newton regularizations*, Numer. Math., 114 (2010), pp. 521–548, <https://doi.org/10.1007/s00211-009-0256-0>.
- [20] H.-P. LIU, D. L. ANDERSON, AND H. KANAMORI, *Velocity dispersion due to anelasticity; implications for seismology and mantle composition*, Geophysical Journal International, 47 (1976), pp. 41–58.
- [21] F. MARGOTTI, *On Inexact Newton Methods for Inverse Problems in Banach Spaces*, PhD thesis, Karlsruhe Institut für Technologie (KIT), 2015, <http://digbib.ubka.uni-karlsruhe.de/volltexte/1000048606>.
- [22] F. MARGOTTI AND A. RIEDER, *An inexact Newton regularization in Banach spaces based on the nonstationary iterated Tikhonov method*, J. Inverse Ill-Posed Probl., 23 (2015), pp. 373–392, <https://doi.org/10.1515/jiip-2014-0035>.
- [23] M. RENARDY AND R. C. ROGERS, *An introduction to partial differential equations*, vol. 13,

- Springer Science & Business Media, 2006.
- [24] A. RIEDER, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, *Inverse Problems*, 15 (1999), pp. 309–327, <https://doi.org/10.1088/0266-5611/15/1/028>.
  - [25] A. RIEDER, *Inexact Newton regularization using conjugate gradients as inner iteration*, *SIAM J. Numer. Anal.*, 43 (2005), pp. 604–622, <https://doi.org/10.1137/040604029>.
  - [26] Y. SAAD, *Iterative methods for sparse linear systems*, vol. 82, SIAM, 2003.
  - [27] P. YANG, R. BROSSIER, L. MÉTIVIER, AND J. VIRIEUX, *A review on the systematic formulation of 3-d multiparameter full waveform inversion in viscoelastic medium*, *Geophysical Journal International*, 207 (2016), pp. 129–149, <http://dx.doi.org/10.1093/gji/ggw262>.
  - [28] P. YANG, R. BROSSIER, L. MÉTIVIER, J. VIRIEUX, AND W. ZHOU, *A time-domain preconditioned truncated Newton approach to visco-acoustic multiparameter full waveform inversion*, *SIAM Journal on Scientific Computing*, 40 (2018), pp. B1101–B1130, <https://doi.org/10.1137/17M1126126>.
  - [29] U. ZELTMANN, *The Viscoelastic Seismic Model: Existence, Uniqueness and Differentiability with Respect to Parameters*, PhD thesis, Karlsruhe Institute of Technology, 2018, <http://dx.doi.org/10.5445/IR/1000093989>.