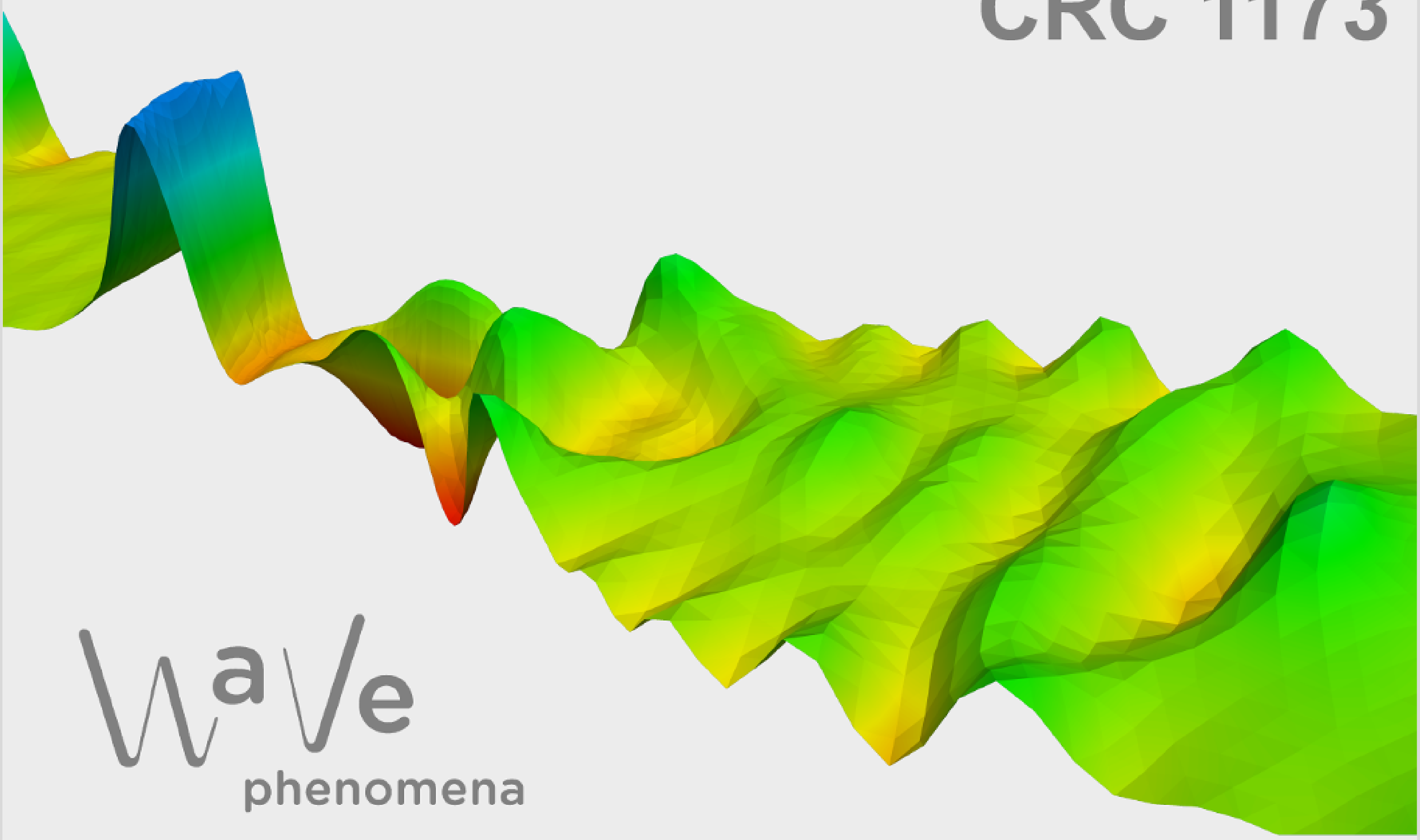
An all-at-once approach to full waveform seismic inversion in the viscoelastic regime

Andreas Rieder

CRC Preprint 2020/16, May 2020

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

TU WIEN

Funded by

DFG

ISSN 2365-662X

AN ALL-AT-ONCE APPROACH TO FULL WAVEFORM SEISMIC INVERSION IN THE VISCOELASTIC REGIME

ANDREAS RIEDER

ABSTRACT. Full waveform seismic inversion (FWI) in the viscoelastic regime entails the task of identifying parameters in the viscoelastic wave equation from partial waveform measurements. Traditionally, one frames this nonlinear problem as an operator equation for the parameter-to-state map. Alternatively, in an all-at-once approach one augments the nonlinear operator by the viscoelastic wave equation as an additional component and considers the states as additional variables. Hence, parameters and states are sought-for simultaneously. In this article, we give a mathematically rigorous all-at-once version of FWI in a functional analytical formulation. Further, the corresponding nonlinear map is shown to be Fréchet differentiable and the adjoint operator of the Fréchet derivative is given in an explicit way suitable for implementation in a Newton-type/gradient-based regularization scheme.

1. Introduction

Parameter identification problems for partial differential equations (pde) are usually formulated as nonlinear operator equations. The involved nonlinear operator is the parameter-to-state map Φ which maps the parameter to the solution (state) of the pde. Applying an iterative regularization scheme for its solution typically requires the evaluation of Φ, Φ' (Fréchet derivative), and Φ'^* (adjoint) at the actual iterate. Each of these evaluations means solving the pde or a related one which is computationally expensive, especially for time-dependent problems like the wave equation.

The same situation appears in control theory and optimization under pde constraints. In these fields emerged quite naturally the idea not to solve the pde but append the constraint to the Lagrangian function and search for its critical points, see, e.g., [8, 9, 16]. This modus operandi is known as the 'all-at-once' approach since one deals simultaneously with the actual optimization task and the underlying pde.

Meanwhile all-at-once methods have diffused into the field of inverse and ill-posed problems. In [11] regularization results have been shown for abstract all-at-once formulations of variational and Newton-type methods. Also implementation issues are discussed. These results have been extended to a family of dynamic inverse problems [12].

To highlight more clearly the two different concepts, let $L = L(p)$ be a differential operator depending on parameters p and let u be the solution of the pde $L(p)u = f$. Assume that we want to recover p from measurements $y = \Psi u$ (Ψ is the measurement

Date: May 25, 2020.

2000 *Mathematics Subject Classification.* 35F10, 35R30, 86A22.

Key words and phrases. all-at-once method, full waveform inversion, wavefield reconstruction inversion, viscoelastic wave equation, adjoint state method, nonlinear inverse and ill-posed problem.

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

The author thanks Andreas Kirsch, Christian Rheinbay, and Roland Schnaubelt for clarifying discussions.

operator). The traditional or reduced approach would be to solve

$$\Phi_{\text{red}}(p) = y$$

where $\Phi_{\text{red}}\colon p \mapsto \Psi L(p)^{-1} f$ is the parameter-to-state map. Evaluation of Φ_{red} means solving the pde. For the all-at-once formulation we define the map $\Phi_{\text{all}}\colon (v,p) \mapsto (L(p)v - f, \Psi v)$. Here we need to solve

$$\Phi_{\text{all}}(u,p) = (0,y).$$

It is the goal of this work to give a mathematically clean all-at-once formulation for full waveform inversion (FWI) in the viscoelastic regime and to provide necessary ingredients to set up Newton/gradient -type regularization schemes. FWI is the up-to-date inversion procedure in geophysical exploration taking full advantage of the amplitude and phase information of seismic recordings, see, e.g., [7, 19]. To unleash the full potential of FWI, it needs to be based on a realistic model for wave propagation in dispersive media. Here we rely on the widely accepted viscoelastic wave equation, see (1) and (4) below.

To the best of our knowledge the first results about an all-at-once version of FWI, termed wavefield reconstruction inversion (WRI), have been reported in [17], however, for the acoustic wave equation in the frequency domain (Helmholtz equation). In the subsequent publication [18] WRI has been generalized to other non-dynamic (frequency domain) inverse problems. We refer to [1] for a recent study of several extensions of WRI (still in the frequency range). All these contributions are done on a discrete, matrix-based level in an optimization, data fitting framework.

The reported advantages of the all-at-once approach to FWI are:

- No solution of the wave and adjoint wave equation is required. Only the differential operator has to be applied to the actual iterate (on the discrete side only matrix-vector products have to be performed).
- The nonlinearity is moderate, meaning that Newton-like solvers for the inverse problem converge fast, that is, within a few iterations, even for a poor starting guess (in the geophysical language: cycle skipping is mitigated). To a certain extent, this observation is supported by the statement of Lemma 3.5.

The remainder of this paper is organized as follows. In the next section we recall the viscoelastic model of wave propagation and write it neatly as an evolution equation in a Hilbert space. It thus fits into the abstract setting of Section 3 where our all-at-once formulation is introduced in a rather general situation. We show its well-definedness, prove Fréchet differentiability (Proposition 4.3), and provide a representation of the adjoint operator of the Fréchet derivative (Proposition 4.4). These are the components of iterative regularization schemes to solve the seismic inverse problem. Finally, in Section 4 we express these results explicitly for the viscoelastic wave equation. Moreover, we show that all-at-once FWI is a locally ill-posed inverse problem just like traditional FWI (Proposition 4.2).

2. Viscoelasticity

The material of this section can already be found in previous publications, see, e.g., [14]. We need to recall it nevertheless for sake of completeness, for introducing some notation, and for indicating a minor flaw, see Remark 2.1 below.

Waves propagating in the earth exhibit damping (loss of energy) which is not reflected by the standard elastic wave equation. Thus, the elastic wave equation has to be augmented by a mechanism which models dispersion and attenuation. Several of these mechanisms are known in the literature which are all closely related, see [7, Chap. 5] and [20, Chap. 2] for an overview and references.

The viscoelastic wave equation in the velocity stress formulation based on the generalized standard linear solid rheology reads: Using $L \in \mathbb{N}$ memory tensors $\boldsymbol{\eta}_l \colon [0,T] \times D \to \mathbb{R}^{3\times 3}_{\mathrm{sym}}$, $l = 1, \dots, L$, the new formulation reads

$$\rho\, \partial_t \mathbf{v} = \operatorname{div} \boldsymbol{\sigma} + \mathbf{f} \qquad \text{in }]0,T[\times D, \tag{1a}$$

$$\partial_t \boldsymbol{\sigma} = C\big((1 + L\tau_{\mathrm{S}})\mu, (1 + L\tau_{\mathrm{P}})\pi\big)\, \varepsilon(\mathbf{v}) + \sum_{l=1}^{L} \boldsymbol{\eta}_l \qquad \text{in }]0,T[\times D, \tag{1b}$$

$$-\tau_{\boldsymbol{\sigma},l} \partial_t \boldsymbol{\eta}_l = C\big(\tau_{\mathrm{S}}\mu, \tau_{\mathrm{P}}\pi\big)\, \varepsilon(\mathbf{v}) + \boldsymbol{\eta}_l, \quad l = 1, \dots, L, \qquad \text{in }]0,T[\times D \tag{1c}$$

where $D \subset \mathbb{R}^3$ is a Lipschitz domain. The functions $\tau_{\mathrm{P}}, \tau_{\mathrm{S}} \colon D \to \mathbb{R}$ are scaling factors for the relaxed moduli π and μ, respectively. They have been introduced by [2].

In (1a), $\mathbf{f}$ denotes the external volume force density and ρ is the mass density. The linear maps $C(m, p)$ in (1b) and (1c) are defined as

$$C(m,p) \colon \mathbb{R}^{3\times 3} \to \mathbb{R}^{3\times 3}, \qquad C(m,p)\mathbf{M} = 2m\mathbf{M} + (p - 2m) \operatorname{tr}(\mathbf{M})\mathbf{I}, \tag{2}$$

for $m, p \in \mathbb{R}$ (C is known as Hooke's tensor). Further, $\mathbf{I} \in \mathbb{R}^{3\times 3}$ is the identity matrix and $\operatorname{tr}(\mathbf{M})$ denotes the trace of $\mathbf{M} \in \mathbb{R}^{3\times 3}$. Finally,

$$\boldsymbol{\varepsilon}(\mathbf{v}) = \frac{1}{2}\big[(\nabla_x \mathbf{v})^\top + \nabla_x \mathbf{v}\big]$$

is the linearized strain rate.

Wave propagation is frequency-dependent and the numbers $\tau_{\boldsymbol{\sigma},l} > 0$, $l = 1, \dots, L$, are used to model this dependency over a frequency band with center frequency ω_0. Within this band the rate of the full energy over the dissipated energy remains nearly constant. This observation lets us determine the stress relaxation times $\tau_{\boldsymbol{\sigma},l}$ by a least-squares approach [3, 4].

Now the frequency-dependent phase velocities of P- and S-waves are given by

$$v_{\mathrm{P}}^2 = \frac{\pi}{\rho}(1 + \tau_{\mathrm{P}}\alpha) \;\text{ and }\; v_{\mathrm{S}}^2 = \frac{\mu}{\rho}(1 + \tau_{\mathrm{S}}\alpha) \;\text{ with } \alpha = \sum_{l=1}^{L} \frac{\omega_0^2 \tau_{\boldsymbol{\sigma},l}^2}{1 + \omega_0^2 \tau_{\boldsymbol{\sigma},l}^2}. \tag{3}$$

Full waveform inversion (FWI) means to reconstruct the five spatially dependent parameters $(\rho, v_{\mathrm{S}}, \tau_{\mathrm{S}}, v_{\mathrm{P}}, \tau_{\mathrm{P}})$ from wavefield measurements.

By the transformation, see [20],

$$\begin{pmatrix} \mathbf{v} \\ \boldsymbol{\sigma}_0 \\ \boldsymbol{\sigma}_1 \\ \vdots \\ \boldsymbol{\sigma}_L \end{pmatrix} := \begin{pmatrix} \mathbf{v} \\ \boldsymbol{\sigma} + \sum_{l=1}^{L} \tau_{\boldsymbol{\sigma},l} \boldsymbol{\eta}_l \\ -\tau_{\boldsymbol{\sigma},1} \boldsymbol{\eta}_1 \\ \vdots \\ -\tau_{\boldsymbol{\sigma},L} \boldsymbol{\eta}_1 \end{pmatrix}$$

we reformulate (1) equivalently into

$$\partial_t \mathbf{v} = \frac{1}{\rho}\operatorname{div}\Big(\sum_{l=0}^{L}\boldsymbol{\sigma}_l\Big) + \frac{1}{\rho}\mathbf{f} \qquad \text{in }]0,T[\times D, \tag{4a}$$

$$\partial_t \boldsymbol{\sigma}_0 = C\big(\mu,\pi\big)\,\varepsilon(\mathbf{v}) \qquad \text{in }]0,T[\times D, \tag{4b}$$

$$\partial_t \boldsymbol{\sigma}_l = C\big(\tau_{\mathrm{S}}\mu, \tau_{\mathrm{P}}\pi\big)\,\varepsilon(\mathbf{v}) - \frac{1}{\tau_{\boldsymbol{\sigma},l}}\boldsymbol{\sigma}_l, \qquad l = 1,\dots,L, \qquad \text{in }]0,T[\times D. \tag{4c}$$

We close the above system by initial conditions

$$\mathbf{v}(0) = \mathbf{v}_0 \ \text{ and } \ \boldsymbol{\sigma}_l(0) = \boldsymbol{\sigma}_{l,0}, \ l = 0,\dots,L. \tag{4d}$$

For a suitable function space[1] X and suitable $w = (\mathbf{w}, \boldsymbol{\psi}_0, \dots, \boldsymbol{\psi}_L) \in X$ we define operators A, B, and Q mapping into X by

$$Aw = -\begin{pmatrix} \operatorname{div}\big(\sum_{l=0}^{L}\boldsymbol{\psi}_l\big) \\ \varepsilon(\mathbf{w}) \\ \vdots \\ \varepsilon(\mathbf{w}) \end{pmatrix}, \quad B^{-1}w = \begin{pmatrix} \frac{1}{\rho}\mathbf{w} \\ C\big(\mu,\pi\big)\boldsymbol{\psi}_0 \\ C\big(\tau_{\mathrm{S}}\mu,\tau_{\mathrm{P}}\pi\big)\boldsymbol{\psi}_1 \\ \vdots \\ C\big(\tau_{\mathrm{S}}\mu,\tau_{\mathrm{P}}\pi\big)\boldsymbol{\psi}_L \end{pmatrix}, \quad Qw = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \frac{1}{\tau_{\boldsymbol{\sigma},1}}\boldsymbol{\psi}_1 \\ \vdots \\ \frac{1}{\tau_{\boldsymbol{\sigma},L}}\boldsymbol{\psi}_L \end{pmatrix}. \tag{5}$$

Now (4) can be formulated as

$$Bu'(t) + Au(t) + BQu(t) = f(t)$$

where $u = (\mathbf{v}, \boldsymbol{\sigma}_0, \dots, \boldsymbol{\sigma}_L)$ and $f = (\mathbf{f}, \mathbf{0}, \dots, \mathbf{0})$.

The operator B depends solely on the five parameters $(\rho, v_{\mathrm{S}}, \tau_{\mathrm{S}}, v_{\mathrm{P}}, \tau_{\mathrm{P}})$ since, according to (3),

$$\pi = \frac{\rho\, v_{\mathrm{P}}^2}{1+\tau_{\mathrm{P}}\alpha} \quad \text{and} \quad \mu = \frac{\rho\, v_{\mathrm{S}}^2}{1+\tau_{\mathrm{S}}\alpha}. \tag{6}$$

Remark 2.1. *Please note that in the previous publication [14] we had wrongly an additional factor L in both arguments of the Hooke tensor C in (1c) and (4c). The affected results can be straightforwardly corrected though.*

3. Abstract framework

3.1. **The setting.** We consider an abstract evolution equation in a Hilbert space X of the form

$$Bu'(t) + Au(t) + BQu(t) = f(t), \quad t \in]0,T[, \quad u(0) = u_0, \tag{7}$$

under the following general hypotheses: $T > 0$, $u_0 \in X$,

B belongs to the Banach space $\mathcal{L}^*(X) = \{P \in \mathcal{L}(X) : P^* = P\}$ and satisfies $\langle Bx, x\rangle_X = \langle x, Bx\rangle_X \geq \beta\|x\|_X^2$ for some $\beta > 0$ and for all $x \in X$,

$A\colon \mathsf{D}(A) \subset X \to X$ is a maximal monotone operator: $\langle Ax, x\rangle_X \geq 0$ for all $x \in X$ and $I + A\colon \mathsf{D}(A) \to X$ is onto (I is the identity, $\mathsf{D}(A)$ is the domain of A),

$Q \in \mathcal{L}(X)$, and $f \in L^1([0,T], X)$.

[1] Details follow below.

In [14] it has been shown that the three operators from (5) are well defined and satisfy our general hypotheses in a precise mathematical setting. Also the viscoacoustic wave equation can be formulated in this abstract setting, see, e.g., [5]. The viscoacoustic equation models wave propagation in media which do not support shear stress.[2]

Under the above general assumptions $B^{-1}A+Q$ generates a C_0-semigroup on $(X, \langle\cdot,\cdot\rangle_B)$ with the weighted inner product $\langle\cdot,\cdot\rangle_B = \langle B\cdot,\cdot\rangle_X$, see [13, 14]. Hence, standard existence and uniqueness results apply, for instance (7) has a unique integrated solution $u \in \mathcal{C}([0,T],X)$ satisfying $\int_0^t u(s)\,\mathrm{d}s \in \mathsf{D}(A)$, $t \in]0,T[$, and

$$Bu(t) + (A+BQ)\int_0^t u(s)\,\mathrm{d}s = Bu_0 + \int_0^t f(s)\,\mathrm{d}s, \quad t \in]0,T[, \tag{8}$$

which coincides with the mild solution, see, e.g., [15, Prop. 2.15].

Next we extend A by $A_-\colon \mathsf{D}(A_-) \subset X_- \to X_-$ to X_- which is the completion[3] of X with respect to the weaker norm $\|\cdot\|_- = \|R(\mu,A)\cdot\|_X$ where $R(\cdot,A)$ denotes the resolvent of A and μ is in $\rho(A)$, the resolvent set of A. Note that a different choice of μ yields an equivalent norm. For instance, under our assumptions on A we could set $\mu = -1$, that is, $R(-1,A) = (I+A)^{-1}$. Further, $\mathsf{D}(A_-) = X$ ($\|\cdot\|_X$ is the graph norm of A_-) and $A_- \in \mathcal{L}(X,X_-)$. See, e.g., [6, Chap. 2.5] and [15, Chap. 2.2] for the details.

Using A_- we generalize (8) slightly to

$$Bu(t) + (A_- +BQ)\int_0^t u(s)\,\mathrm{d}s = Bu_0 + \int_0^t f(s)\,\mathrm{d}s, \quad t \in]0,T[. \tag{9}$$

Obviously, the integrated solution of (7) that is, the solution of (8), solves (9). Conversely, we have the following result.

Lemma 3.1. *If $u \in \mathcal{C}([0,T],X)$ solves (9) then $Bu \in \mathcal{C}^1([0,T],X_-)$. Further, we have that $u(0) = u_0$ and*

$$(Bu)'(t) + (A_- + BQ)u(t) = f(t), \quad t \in]0,T[, \tag{10}$$

in the weaker space X_-. If, additionally, B can be extended to X_- continuously invertible then $u \in \mathcal{C}^1([0,T],X_-)$ and $(Bu)'(t) = Bu'(t)$.

Proof. From (9) we get for $t \in]0,T[$ and $0 < |h|$ sufficiently small that

$$\frac{1}{h}\big(Bu(t+h) - Bu(t)\big) = -(A_- + BQ)\frac{1}{h}\int_t^{t+h} u(s)\,\mathrm{d}s + \frac{1}{h}\int_t^{t+h} f(s)\,\mathrm{d}s.$$

The terms on the right hands side converge in X_- when $h \to 0$. Indeed, $\frac{1}{h}\int_t^{t+h} f(s)\,\mathrm{d}s$ and $BQ\frac{1}{h}\int_t^{t+h} u(s)\,\mathrm{d}s$ converge even in X to $f(t)$ and $BQu(t)$, respectively. Finally, since $A_- \in \mathcal{L}(X,X_-)$ the limit of $A_-\frac{1}{h}\int_t^{t+h} u(s)\,\mathrm{d}s$ exists in X_- and is equal to $A_-u(t)$. $\square$

Remark 3.2. *In this remark we explain why we switch from (7) to (9). In general, the mild solution of (7) is not differentiable in time and lies not in the domain of A. Hence, it does not satisfy the differential equation (7). This is why we consider the integrated version (8). Moreover, we introduced (9) with the extension A_- because – as a bounded operator from X to X_- – it is Fréchet differentiable unlike the unbounded $A\colon \mathsf{D}(A) \subset X \to X$.*

[2]The viscoacoustic equation can be derived formally from (4) by setting the shear modulus $\mu = 0$, defining the new state variables $p_l = \mathrm{tr}(\boldsymbol{\sigma}_l)$, $l = 0,\ldots,L$, and taking the traces of (4b) and (4c).

[3]The standard notation for X_- is X_{-1} (as an element of a Sobolev tower/scale [6, Chap. 2.5]).

3.2. **Abstract all-at-once formulation.** We want to formulate FWI as a nonlinear operator equation. To this end we first introduce some abbreviations: Let $\mathcal{H} = L^2([0,T], X)$, $\mathcal{H}_- = L^2([0,T], X_-)$, and let the linear operator $\Psi\colon \mathcal{H} \to \mathbb{R}^N$, $N \in \mathbb{N}$, model the measurement/sampling process (the image of Ψ is the space of seismograms in the geophysical application). Moreover, let $J \in \mathcal{L}(\mathcal{H})$ denote the integration operator: $Jv(t) = \int_0^t v(s)\mathrm{d}s$. Note that $JK = KJ$ for any $K \in \mathcal{L}(X)$.

Now we define the following map related to (9):

$$\begin{aligned} F\colon \mathcal{H} \times \mathcal{B} \subset \mathcal{H} \times \mathcal{L}^*(X) &\to \mathcal{H}_- \times \mathbb{R}^N, \\ (v,P)^\top &\mapsto \big(P(v-u_0) + (A_- + PQ)Jv - Jf,\ \Psi v\big)^\top \end{aligned} \tag{11}$$

where

$$\mathcal{B} = \big\{B \in \mathcal{L}^*(X) : \beta_- \|x\|_X^2 \le \langle Bx, x\rangle_X \le \beta_+ \|x\|_X^2\big\} \tag{12}$$

for given $0 < \beta_- < \beta_+ < \infty$.

The following lemma explains our definition of F.

Lemma 3.3. *If $(u,B)^\top \in \mathcal{H} \times \mathcal{B}$ satisfies $F(u,B) = (0,\Sigma)^\top$ for a given $\Sigma \in \mathbb{R}^N$ then $u \in \mathcal{C}([0,T], X)$ solves (10) in X_- (and generates the seismogram Σ).*

Proof. Obviously, $u \in L^2([0,T], X)$ satisfies the equation of (9) a.e. in $[0,T]$. From $u(t) = u_0 + \int_0^t B^{-1} f(s)\,\mathrm{d}s - (B^{-1}A_- + Q)\int_0^t u(s)\,\mathrm{d}s$ we deduce that u is continuous. Hence, Lemma 3.1 applies. □

Now, FWI can be phrased in the following way: Given $\Sigma \in \mathbb{R}^N$ (a seismogram) find a pair $(u,B)^\top \in \mathcal{H} \times \mathcal{B}$ such that

$$F(u,B) = \begin{pmatrix} 0 \\ \Sigma \end{pmatrix}.$$

Remark 3.4. *Traditional (or reduced) FWI can be formulated as the operator equation $F_{\text{red}}(B) = \Sigma$ with the parameter-to-state map $F_{\text{red}}\colon B \mapsto \Psi u$ where $u \in \mathcal{H}$ solves (7) with respect to $B \in \mathcal{B}$.*

It will prove convenient to split F into $F = F_1 + F_2 + F_3$ with

$$F_1(v,P) = \begin{pmatrix} P(I+QJ)v \\ 0 \end{pmatrix}, \quad F_2(v,P) = \begin{pmatrix} A_- Jv - Pu_0 \\ \Psi v \end{pmatrix}, \quad F_3(v,P) = -\begin{pmatrix} Jf \\ 0 \end{pmatrix}.$$

Please observe that $F_1\colon \mathcal{H} \times \mathcal{L}^*(X) \to \mathcal{H} \times \mathbb{R}^N$ is bilinear and bounded, $F_2 \in \mathcal{L}(\mathcal{H} \times \mathcal{L}^*(X), \mathcal{H}_- \times \mathbb{R}^N)$, and F_3 is constant. Hence, F has simple Fréchet derivatives.

Lemma 3.5. *The mapping F from (11) is Fréchet differentiable at any interior point $(v,P) \in \mathcal{H} \times \mathcal{B}$ with*

$$\begin{aligned} F'(v,P)\begin{bmatrix} \widehat{v} \\ \widehat{P} \end{bmatrix} &= F_1(\widehat{v}, P) + F_1(v, \widehat{P}) + F_2(\widehat{v}, \widehat{P}) \\ &= \begin{pmatrix} P(I+QJ)\widehat{v} + \widehat{P}(I+QJ)v + A_- J\widehat{v} - \widehat{P}u_0 \\ \Psi\widehat{v} \end{pmatrix}. \end{aligned}$$

The second derivative is given by

$$F''(v,P)\left[\begin{pmatrix}\widehat{v}_1\\ \widehat{P}_1\end{pmatrix},\begin{pmatrix}\widehat{v}_2\\ \widehat{P}_2\end{pmatrix}\right]=F_1(\widehat{v}_1,\widehat{P}_2)+F_1(\widehat{v}_2,\widehat{P}_1).$$

All higher derivatives vanish identically.

Finding an explicit representation of the adjoint $F'(v,P)^*\colon \mathcal{H}'_- \times \mathbb{R}^N \to \mathcal{H}' \times \mathcal{L}^*(X)'$ poses no challenge.

Lemma 3.6. *For any interior point $(v,P) \in \mathcal{H} \times \mathcal{B}$ we have that*

$$F'(v,P)^*\begin{pmatrix} g\\ \Sigma\end{pmatrix}=\big((I+J^*Q^*)Pg+J^*A_-^*g+\Psi^*\Sigma,\ \ell_{v,g}\big)$$

where $\ell_{v,g} \in \mathcal{L}^*(X)'$ is the functional $\ell_{v,g}(\widehat{P}) = \langle \widehat{P}g,(I+QJ)v-u_0\rangle_{\mathcal{H}}$ and $J^*w(t) = \int_t^T w(s)\mathrm{d}s$. Note that $\widehat{P}g$ is well defined for $\widehat{P} \in \mathcal{L}(X)$ since $g \in \mathcal{H}'_- \subset \mathcal{H}' \simeq \mathcal{H}$.

Proof. We have

$$\left[F'(v,P)^*\begin{pmatrix} g\\ \Sigma\end{pmatrix}\right]\begin{pmatrix}\widehat{v}\\ \widehat{P}\end{pmatrix}=\left\langle\begin{pmatrix} g\\ \Sigma\end{pmatrix},F'(v,P)\begin{pmatrix}\widehat{v}\\ \widehat{P}\end{pmatrix}\right\rangle_{(\mathcal{H}'_-\times\mathbb{R}^N)\times(\mathcal{H}_-\times\mathbb{R}^N)}\tag{13}$$

$$=\big\langle g,P(I+QJ)\widehat{v}+A_-J\widehat{v}\big\rangle_{\mathcal{H}}+\big\langle g,\widehat{P}(I+QJ)v-\widehat{P}u_0\big\rangle_{\mathcal{H}}+\big\langle\Sigma,\Psi\widehat{v}\big\rangle_{\mathbb{R}^N}$$

from which the assertion follows immediately. □

4. Application to full waveform inversion

4.1. **The setting and basic definitions.** We recall the basic concepts from [14].

The Hilbert space underlying the viscoelastic wave equation (4) is

$$X = L^2(D,\mathbb{R}^3)\times L^2(D,\mathbb{R}^{3\times3}_{\text{sym}})^{1+L}$$

with inner product

$$\big\langle(\mathbf{v},\boldsymbol{\sigma}_0,\dots,\boldsymbol{\sigma}_L),(\mathbf{w},\boldsymbol{\psi}_0,\dots,\boldsymbol{\psi}_l)\big\rangle_X=\int_D\Big(\mathbf{v}\cdot\mathbf{w}+\sum_{l=0}^L\boldsymbol{\sigma}_l:\boldsymbol{\psi}_l\Big)\,\mathrm{d}x.$$

The colon represents the Frobenius inner product of matrices.

Let the boundary ∂D of the bounded Lipschitz domain D be split into disjoint parts $\partial D = \partial D_D \dot\cup \partial D_N$ and let $\mathbf{n}$ be the outer normal vector on ∂D_N. Then, we set

$$\mathsf{D}(A)=\Big\{(\mathbf{w},\boldsymbol{\psi}_0,\dots\boldsymbol{\psi}_L)\in H^1_D\times H(\operatorname{div})^{1+L}:\sum_{l=0}^L\boldsymbol{\psi}_l\mathbf{n}=0\text{ on }\partial D_N\Big\}$$

for the domain of A from (5). Here, $H^1_D = \{\mathbf{v} \in H^1(D,\mathbb{R}^3) : \mathbf{v} = 0 \text{ on } \partial D_D\}$ and $H(\operatorname{div}) = \big\{\boldsymbol{\sigma} \in L^2\big(D,\mathbb{R}^{3\times3}_{\text{sym}}\big) : \operatorname{div}\big(\sum_{l=0}^L \boldsymbol{\sigma}_l\big) \in L^2(D)\big\}$. The following result is validated in [14, Lem. 4.1].

Lemma 4.1. *The operator A as defined in (5) with $\mathsf{D}(A) \subset X$ is maximal monotone and skew-symmetric, i.e., $A^* = -A$.*

Next we present the representation of B for the viscoelastic setting. A crucial ingredient is the operator C of (2) which maps $\mathsf{D}(C) = \big\{(m,p) \in \mathbb{R}^2 : \underline{\mathrm{m}} \le m \le \overline{\mathrm{m}},\ \underline{\mathrm{p}} \le p \le \overline{\mathrm{p}}\big\}$

into $\mathrm{Aut}(\mathbb{R}^{3\times 3}_{\mathrm{sym}})$[4] with constants $0 < \underline{\mathrm{m}} < \overline{\mathrm{m}}$ and $0 < \underline{\mathrm{p}} < \overline{\mathrm{p}}$ such that $3\underline{\mathrm{p}} > 4\overline{\mathrm{m}}$. For $(m,p) \in \mathsf{D}(C)$,

$$\widetilde{C}(m,p) := C(m,p)^{-1} = C\left(\frac{1}{4m}, \frac{p-m}{m(3p-4m)}\right). \tag{14}$$

We have

$$C(m,p)\mathbf{M} : \mathbf{N} = \mathbf{M} : C(m,p)\mathbf{N} \tag{15}$$

and

$$\min\{2\underline{\mathrm{m}}, 3\underline{\mathrm{p}} - 4\overline{\mathrm{m}}\}\, \mathbf{M} : \mathbf{M} \le C(m,p)\mathbf{M} : \mathbf{M} \le \max\{2\overline{\mathrm{m}}, 3\overline{\mathrm{p}} - 4\underline{\mathrm{m}}\}\, \mathbf{M} : \mathbf{M},$$

see, e.g., [20, Lemma 50]. If $\rho(x) > 0$, $\big(\mu_0(x), \pi_0(x)\big), \big(\tau_{\mathrm{S}}(x)\mu_0(x), \tau_{\mathrm{P}}(x)\pi_0(x)\big) \in \mathsf{D}(C)$ for almost all $x \in D$ then

$$B\begin{pmatrix} \mathbf{w} \\ \boldsymbol{\psi}_0 \\ \boldsymbol{\psi}_1 \\ \vdots \\ \boldsymbol{\psi}_L \end{pmatrix} = \begin{pmatrix} \rho\,\mathbf{w} \\ \widetilde{C}(\mu,\pi)\boldsymbol{\psi}_0 \\ \widetilde{C}(\tau_{\mathrm{S}}\mu, \tau_{\mathrm{P}}\pi)\boldsymbol{\psi}_1 \\ \vdots \\ \widetilde{C}(\tau_{\mathrm{S}}\mu, \tau_{\mathrm{P}}\pi)\boldsymbol{\psi}_L \end{pmatrix} \tag{16}$$

yields a uniformly positive $B \in \mathcal{L}^*(X)$ as required by the general hypotheses at the beginning of the former section. We conclude that the abstract all-at-once formulation of Section 3 is well defined for the viscoelastic wave equation (4) provided the initial values (4d) are in X.

4.2. **All-at-once full waveform operator.** In FWI one wants to reconstruct the five parameters $\mathbf{p} = (\rho, v_{\mathrm{S}}, \tau_{\mathrm{S}}, v_{\mathrm{P}}, \tau_{\mathrm{P}})$ from observed wavefields. Therefore we define here an all-at-once operator $\Phi(w, \mathbf{p}) := F(w, V(\mathbf{p}))$ where $V\colon \mathbf{p} \mapsto B$ and F is the mapping from (11).

A physically meaningful domain of definition for V is

$$\begin{aligned}\mathsf{D}(V) = \big\{(\rho, v_{\mathrm{S}}, \tau_{\mathrm{S}}, v_{\mathrm{P}}, \tau_{\mathrm{P}}) \in L^\infty(D)^5 : \rho_{\min} \le \rho(\cdot) \le \rho_{\max},\ v_{\mathrm{P},\min} \le v_{\mathrm{P}}(\cdot) \le v_{\mathrm{P},\max}, \\ v_{\mathrm{S},\min} \le v_{\mathrm{S}}(\cdot) \le v_{\mathrm{S},\max},\ \tau_{\mathrm{P},\min} \le \tau_{\mathrm{P}}(\cdot) \le \tau_{\mathrm{P},\max},\ \tau_{\mathrm{S},\min} \le \tau_{\mathrm{S}}(\cdot) \le \tau_{\mathrm{S},\max} \text{ a.e. in } D\big\}\end{aligned}$$

with suitable positive bounds $0 < \rho_{\min} < \rho_{\max} < \infty$, etc. Note that $\underline{\mathrm{p}}$, $\overline{\mathrm{p}}$, $\underline{\mathrm{m}}$, and $\overline{\mathrm{m}}$ can be defined in terms of $\rho_{\min}, \rho_{\max}, v_{\mathrm{P},\min}$, etc. such that $(\mu,\pi), (\tau_{\mathrm{S}}\mu, \tau_{\mathrm{P}}\pi)$ as functions of $(\rho, v_{\mathrm{P}}, v_{\mathrm{S}}, \tau_{\mathrm{P}}, \tau_{\mathrm{S}}) \in \mathsf{D}(V)$ are in $\mathsf{D}(C)$, see [14].

Hence, we have a well-defined mapping

$$V\colon \mathsf{D}(V) \subset L^\infty(D)^5 \to \mathcal{B} \subset \mathcal{L}^*(X), \quad (\rho, v_{\mathrm{S}}, \tau_{\mathrm{S}}, v_{\mathrm{P}}, \tau_{\mathrm{P}}) \mapsto B,$$

where B is given in (16) via (6). We indeed have that $V(\mathsf{D}(V)) \subset \mathcal{B}$ when β_- and β_+ are chosen appropriately in (12).

Finally, we get the *all-at-once full waveform forward operator* by

$$\begin{gathered}\Phi\colon \mathsf{D}(\Phi) \subset \mathcal{H} \times L^\infty(D)^5 \to \mathcal{H}_- \times \mathbb{R}^N, \\ \big(w, (\rho, v_{\mathrm{S}}, \tau_{\mathrm{S}}, v_{\mathrm{P}}, \tau_{\mathrm{P}})\big) \mapsto F\big(w, V(\rho, v_{\mathrm{S}}, \tau_{\mathrm{S}}, v_{\mathrm{P}}, \tau_{\mathrm{P}})\big),\end{gathered}$$

[4]This is the space of linear maps from $\mathbb{R}^{3\times 3}_{\mathrm{sym}}$ into itself (space of automorphisms).

with $\mathsf{D}(\Phi) = \mathcal{H} \times \mathsf{D}(V)$. In this setting FWI means: Given a seismogram $\Sigma \in \mathbb{R}^N$ find a wavefield $u \in L^2\big([0,T], X\big)$ and a set of parameters $\mathbf{p} = (\rho, v_\mathrm{S}, \tau_\mathrm{S}, v_\mathrm{P}, \tau_\mathrm{P}) \in \mathsf{D}(V)$ such that

$$\Phi(u, \mathbf{p}) = \begin{pmatrix} 0 \\ \Sigma \end{pmatrix}. \tag{17}$$

If $(u, \mathbf{p})^\top$ solves the above equation then $u = (\mathbf{v}, \boldsymbol{\sigma}_0, \ldots, \boldsymbol{\sigma}_L)^\top$ is the mild solution of (4) with respect to the parameter set given by $\mathbf{p}$ and generates the seismogram Σ (Lemma 3.3).

Just like the reduced FWI formulation, see [14, Theorem 4.3], the all-at-once version (17) is locally ill-posed. Local ill-posedness of an inverse problem was introduced in [10]: let $\Theta\colon \mathsf{D}(\Theta) \subset V \to W$ operate between infinite-dimensional normed spaces. We say that $\Theta(\cdot) = w$ is locally ill-posed at $v^+ \in \mathsf{D}(\Theta)$ satisfying $\Theta(v^+) = w$ if in any neighborhood O of v^+ we can find a sequence $\{\eta_k\} \subset O \cap \mathsf{D}(\Theta)$ which does not converge to v^+, yet, we have $\Theta(\eta_k) \to w$.

Proposition 4.2. *The inverse problem $\Phi(\cdot,\cdot) = (0, \Sigma)^\top$ is locally ill-posed at any interior point $(u, \mathbf{p})$ of $\mathsf{D}(\Phi)$.*

Proof. We adapt the proof of Theorem 4.3 of [14] to the present situation. Choose a $x_o \in D$ and define the balls $K_k = \{x \in \mathbb{R}^3 : |x - x_o| \leq \epsilon/k\}$, $k \in \mathbb{N}$, where $\epsilon > 0$ is small enough to guarantee $K_k \subset D$. Set $\mathbf{p}_k := \mathbf{p} + r(\chi_k, \chi_k, \chi_k, \chi_k, \chi_k)$ where $r > 0$ and χ_k is the indicator function of K_k. For r sufficiently small we have $\mathbf{p}_k \in \mathsf{D}(V)$ and $\|\mathbf{p}_k - \mathbf{p}\|_{L^\infty(D)^5} = r$. So, $\{\mathbf{p}_k\}$ is as close to $\mathbf{p}$ as we wish without converging to it.

By u_k and u let us denote the unique mild solutions of (4) with respect to the parameters in $\mathbf{p}_k$ and $\mathbf{p}$, respectively. Then, $\{(u_k, \mathbf{p}_k)\} \subset \mathsf{D}(\Phi)$, $\{(u, \mathbf{p})\} \subset \mathsf{D}(\Phi)$, $\Phi_1(u_k, \mathbf{p}_k) = 0 = \Phi_1(u, \mathbf{p})$, and $u_k \to u$ in $\mathcal{H}$. The convergence can be verified by showing first that $u - u_k$ is the mild solution of of a related evolution equation and then by applying the stability estimate for mild solutions, see, e.g., [14] for details.

Thus, for all k sufficiently large, $\{(u_k, \mathbf{p}_k)\}$ is as close to $(u, \mathbf{p})$ in $\mathcal{H} \times L^\infty(D)^5$ as we wish without converging to it. However,

$$\|\Phi(u_k, \mathbf{p}_k) - \Phi(u, \mathbf{p})\|_{\mathcal{H}_- \times \mathbb{R}^N} = \left\| \begin{pmatrix} 0 \\ \Psi(u_k - u) \end{pmatrix} \right\|_{\mathcal{H}_- \times \mathbb{R}^N} = \|\Psi(u_k - u)\|_{\mathbb{R}^N} \to 0$$

by continuity of the measurement operator Ψ. □

According to the above result, solving (17) requires regularization. Using Newton-like regularization schemes one needs to implement the Fréchet derivative and its adjoint. In the remainder of this section we provide rather explicit analytic expressions for both.

We obtain the Fréchet derivative of Φ by the chain rule, Lemma 3.5, and the derivative of V which was presented in [14]. The formulation of V' relies on the derivative of $\widetilde{C}$ which is

$$\widetilde{C}'(m,p)\begin{bmatrix} \widehat{m} \\ \widehat{p} \end{bmatrix} = -\widetilde{C}(m,p) \circ C(\widehat{m}, \widehat{p}) \circ \widetilde{C}(m,p) \tag{18}$$

for $(m,p) \in \operatorname{int}(\mathsf{D}(C))$ and $(\widehat{m}, \widehat{p}) \in \mathbb{R}^2$.

Let $\mathbf{p} = (\rho, v_{\rm S}, \tau_{\rm S}, v_{\rm P}, \tau_{\rm P}) \in \operatorname{int}(\mathsf{D}(\Phi))$ and $\widehat{\mathbf{p}} = (\widehat{\rho}, \widehat{v}_{\rm S}, \widehat{\tau}_{\rm S}, \widehat{v}_{\rm P}, \widehat{\tau}_{\rm P}) \in L^\infty(D)^5$. Then, $V'(\mathbf{p})\widehat{\mathbf{p}} \in \mathcal{L}^*(X)$ is given by

$$V'(\mathbf{p})\widehat{\mathbf{p}}\begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}_0\\ \vdots\\ \boldsymbol{\psi}_L\end{pmatrix} = \begin{pmatrix} \widehat{\rho}\,\mathbf{w}\\ -\frac{\widehat{\rho}}{\rho}\widetilde{C}(\mu,\pi)\boldsymbol{\psi}_0 + \rho\widetilde{C}'(\mu,\pi)\begin{bmatrix}\widetilde{\mu}\\ \widetilde{\pi}\end{bmatrix}\boldsymbol{\psi}_0\\ -\frac{\widehat{\rho}}{\rho}\widetilde{C}(\tau_{\rm S}\mu,\tau_{\rm P}\pi)\boldsymbol{\psi}_1 + \rho\,\widetilde{C}'(\tau_{\rm S}\mu,\tau_{\rm P}\pi)\begin{bmatrix}\widehat{\mu}\\ \widehat{\pi}\end{bmatrix}\boldsymbol{\psi}_1\\ \vdots\\ -\frac{\widehat{\rho}}{\rho}\widetilde{C}(\tau_{\rm S}\mu,\tau_{\rm P}\pi)\boldsymbol{\psi}_L + \rho\,\widetilde{C}'(\tau_{\rm S}\mu,\tau_{\rm P}\pi)\begin{bmatrix}\widehat{\mu}\\ \widehat{\pi}\end{bmatrix}\boldsymbol{\psi}_L\end{pmatrix} \tag{19}$$

where μ and π are from (6) and

$$\widetilde{\mu} = a_{\rm S}\,\widehat{v}_{\rm S} - \alpha\, b_{\rm S}\,\widehat{\tau}_{\rm S}, \qquad \widetilde{\pi} = a_{\rm P}\,\widehat{v}_{\rm P} - \alpha\, b_{\rm P}\,\widehat{\tau}_{\rm P}, \tag{20}$$

$$\widehat{\mu} = \tau_{\rm S}\, a_{\rm S}\,\widehat{v}_{\rm S} + b_{\rm S}\,\widehat{\tau}_{\rm S}, \qquad \widehat{\pi} = \tau_{\rm P}\, a_{\rm P}\,\widehat{v}_{\rm P} + b_{\rm P}\,\widehat{\tau}_{\rm P}, \tag{21}$$

with

$$a_{\rm S} = \frac{2v_{\rm S}}{1+\tau_{\rm S}\alpha},\quad b_{\rm S} = \frac{v_{\rm S}^2}{(1+\tau_{\rm S}\alpha)^2},\quad a_{\rm P} = \frac{2v_{\rm P}}{1+\tau_{\rm P}\alpha},\quad b_{\rm P} = \frac{v_{\rm P}^2}{(1+\tau_{\rm P}\alpha)^2}.$$

We introduce new symbolic notation:

$$w^\uparrow(t) := Jw(t) = \int_0^t w(s)\mathrm{d}s \quad\text{and}\quad w^\downarrow(t) := J^*w(t) = \int_t^T w(s)\mathrm{d}s.$$

Proposition 4.3. *Under the assumptions of this section the all-at-once full waveform forward operator Φ is Fréchet differentiable at any interior point $(w,\mathbf{p})$ of $\mathsf{D}(\Phi)$, $w = (\mathbf{w},\boldsymbol{\psi}_0,\dots,\boldsymbol{\psi}_L)$, $\mathbf{p} = (\rho, v_S, \tau_S, v_P, \tau_P)$:*

For $\widehat{w} = (\widehat{\mathbf{w}},\widehat{\boldsymbol{\psi}}_0,\dots,\widehat{\boldsymbol{\psi}}_L) \in \mathcal{H}$ and $\widehat{\mathbf{p}} = (\widehat{\rho},\widehat{v}_S,\widehat{\tau}_S,\widehat{v}_P,\widehat{\tau}_P) \in L^\infty(D)^5$ we have

$$\Phi'(w,\mathbf{p})\begin{bmatrix}\widehat{w}\\ \widehat{\mathbf{p}}\end{bmatrix} =$$

$$\begin{pmatrix} \rho\,\widehat{\mathbf{w}} + \widehat{\rho}\,(\mathbf{w}-\mathbf{v}_0) - \operatorname{div}_-\Big(\sum_{l=0}^L \widehat{\boldsymbol{\psi}}_l^\uparrow\Big)\\ \widetilde{C}(\mu,\pi)\big(\widehat{\boldsymbol{\psi}}_0 - \frac{\widehat{\rho}}{\rho}(\boldsymbol{\psi}_0 - \boldsymbol{\sigma}_{0,0})\big) + \rho\widetilde{C}'(\mu,\pi)\begin{bmatrix}\widetilde{\mu}\\ \widetilde{\pi}\end{bmatrix}(\boldsymbol{\psi}_0 - \boldsymbol{\sigma}_{0,0}) - \varepsilon_-(\widehat{\mathbf{w}}^\uparrow)\\ \widetilde{C}(\tau_S\mu,\tau_P\pi)\big((1+\frac{1}{\tau_{\boldsymbol{\sigma},1}})\widehat{\boldsymbol{\psi}}_1^\uparrow - \frac{\widehat{\rho}}{\rho}(\boldsymbol{\psi}_1 - \boldsymbol{\sigma}_{1,0} + \frac{1}{\tau_{\boldsymbol{\sigma},1}}\boldsymbol{\psi}_1^\uparrow)\big)\\ \qquad +\rho\,\widetilde{C}'(\tau_S\mu,\tau_P\pi)\begin{bmatrix}\widehat{\mu}\\ \widehat{\pi}\end{bmatrix}\big(\boldsymbol{\psi}_1 - \boldsymbol{\sigma}_{1,0} + \frac{1}{\tau_{\boldsymbol{\sigma},1}}\boldsymbol{\psi}_1^\uparrow\big) - \varepsilon_-(\widehat{\mathbf{w}}^\uparrow)\\ \vdots\\ \widetilde{C}(\tau_S\mu,\tau_P\pi)\big((1+\frac{1}{\tau_{\boldsymbol{\sigma},L}})\widehat{\boldsymbol{\psi}}_L^\uparrow - \frac{\widehat{\rho}}{\rho}(\boldsymbol{\psi}_L - \boldsymbol{\sigma}_{L,0} + \frac{1}{\tau_{\boldsymbol{\sigma},L}}\boldsymbol{\psi}_L^\uparrow)\big)\\ \qquad +\rho\,\widetilde{C}'(\tau_S\mu,\tau_P\pi)\begin{bmatrix}\widehat{\mu}\\ \widehat{\pi}\end{bmatrix}\big(\boldsymbol{\psi}_L - \boldsymbol{\sigma}_{L,0} + \frac{1}{\tau_{\boldsymbol{\sigma},L}}\boldsymbol{\psi}_L^\uparrow\big) - \varepsilon_-(\widehat{\mathbf{w}}^\uparrow)\\ \Psi(\widehat{\mathbf{w}},\widehat{\boldsymbol{\psi}}_0,\dots,\widehat{\boldsymbol{\psi}}_L)\end{pmatrix}$$

where div_- *and* ε_- *are the components of* A_-.[5] *Further,* $\mathbf{v}_0$ *and* $\boldsymbol{\sigma}_{l,0}$*,* $l = 0, \dots, L$ *are the initial values, see* (4d)*.*

Proof. The stated expression for Φ' follows readily from the chain rule and Lemma 3.5:

$$\Phi'(w,\mathbf{p})\begin{bmatrix}\widehat{w}\\ \widehat{\mathbf{p}}\end{bmatrix} = F'\big(w, V(\mathbf{p})\big)\begin{bmatrix}\widehat{w}\\ V'(\mathbf{p})\widehat{\mathbf{p}}\end{bmatrix}$$
$$= \begin{pmatrix} V(\mathbf{p})(I+QJ)\widehat{w} + V'(\mathbf{p})\widehat{\mathbf{p}}(I+QJ)w + A_- J\widehat{w} - V'(\mathbf{p})\widehat{\mathbf{p}}u_0 \\ \Psi\widehat{w}\end{pmatrix}.$$

Finally, we plug in the expressions from (5), (16), and (19). □

Proposition 4.4. *The notation and the assumptions are as in the previous proposition. Then, the adjoint* $\Phi'(w,\mathbf{p})^* \in \mathcal{L}\big(\mathcal{H}'_- \times \mathbb{R}^N, \mathcal{H}' \times (L^\infty(D)^5)'\big)$ *is given by*

$$\Phi'(w,\mathbf{p})^*\begin{pmatrix} g\\ \Sigma\end{pmatrix} = \left(\Phi'(w,\mathbf{p})_1^*\begin{pmatrix} g\\ \Sigma\end{pmatrix}, \Phi'(w,\mathbf{p})_2^*\begin{pmatrix} g\\ \Sigma\end{pmatrix}\right)$$

where $g = (\mathbf{g}_{-1}, \mathbf{g}_0, \dots, \mathbf{g}_L) \in L^2\big([0,T], L^2(D,\mathbb{R}^3) \times L^2(D, \mathbb{R}^{3\times 3}_{\text{sym}})^{1+L}\big)$*,* $\Sigma \in \mathbb{R}^N$*,*

$$\Phi'(w,\mathbf{p})_1^*\begin{pmatrix} g\\ \Sigma\end{pmatrix} = \begin{pmatrix} \rho\,\mathbf{g}_{-1} + \operatorname{div}_-\Big(\sum_{l=0}^L \mathbf{g}_l^{\downarrow}\Big) \\ \widetilde{C}(\mu,\pi)\mathbf{g}_0 + \varepsilon(\mathbf{g}_{-1}^{\downarrow}) \\ \widetilde{C}(\tau_S\mu, \tau_P\pi)\big(\mathbf{g}_1 + \frac{1}{\tau_{\boldsymbol{\sigma},1}}\mathbf{g}_1^{\downarrow}\big) + \varepsilon(\mathbf{g}_{-1}^{\downarrow}) \\ \vdots \\ \widetilde{C}(\tau_S\mu, \tau_P\pi)\big(\mathbf{g}_L + \frac{1}{\tau_{\boldsymbol{\sigma},L}}\mathbf{g}_L^{\downarrow}\big) + \varepsilon(\mathbf{g}_{-1}^{\downarrow})\end{pmatrix}^{\top} + \Psi^*\Sigma$$

and

$$\Phi'(w,\mathbf{p})_2^*\begin{pmatrix} g\\ \Sigma\end{pmatrix} = \begin{pmatrix} \int_0^T \Big((\mathbf{w}-\mathbf{v}_0)\cdot\mathbf{g}_{-1} - \frac{1}{\rho}\boldsymbol{\Psi}_0 : \mathbf{g}_0 - \frac{1}{\rho}\sum_{l=1}^L \boldsymbol{\Psi}_l : \mathbf{g}_l\Big)\mathrm{d}t \\ -2a_S\,\rho\int_0^T \big(\boldsymbol{\Psi}_0 \,\triangle\, \mathbf{G}_0 + \tau_S\sum_{l=1}^L \boldsymbol{\Psi}_l \,\triangle\, \mathbf{G}_l\big)\mathrm{d}t \\ 2b_S\,\rho\int_0^T \big(\alpha\boldsymbol{\Psi}_0 \,\triangle\, \mathbf{G}_0 - \sum_{l=1}^L \boldsymbol{\Psi}_l \,\triangle\, \mathbf{G}_l\big)\mathrm{d}t \\ -a_P\,\rho\int_0^T \big(\boldsymbol{\Psi}_0 \star \mathbf{G}_0 + \tau_P\sum_{l=1}^L \boldsymbol{\Psi}_l \star \mathbf{G}_l\big)\mathrm{d}t \\ b_P\,\rho\int_0^T \big(\alpha\boldsymbol{\Psi}_0 \star \mathbf{G}_0 - \sum_{l=1}^L \boldsymbol{\Psi}_l \star \mathbf{G}_l\big)\mathrm{d}t \end{pmatrix}^{\top}$$

with the following abbreviations

(22a) $\boldsymbol{\Psi}_0 = \widetilde{C}(\mu,\pi)(\boldsymbol{\psi}_0 - \boldsymbol{\sigma}_{0,0})$, $\boldsymbol{\Psi}_l = \widetilde{C}(\tau_S\mu, \tau_P\pi)(\boldsymbol{\psi}_l - \boldsymbol{\sigma}_{l,0} + \boldsymbol{\psi}_l^{\uparrow}/\tau_{\boldsymbol{\sigma},l})$, $l = 1, \dots, L$,

(22b) $\mathbf{G}_0 = \widetilde{C}(\mu,\pi)\mathbf{g}_0$, $\mathbf{G}_l = \widetilde{C}(\tau_S\mu, \tau_P\pi)\mathbf{g}_l$, $l = 1, \dots, L$,

and

$$\boldsymbol{\Psi}_l \star \mathbf{G}_l = \operatorname{tr}(\boldsymbol{\Psi}_l)\operatorname{tr}(\mathbf{G}_l), \quad \boldsymbol{\Psi}_l \,\triangle\, \mathbf{G}_l = \boldsymbol{\Psi}_l : \mathbf{G}_l - \boldsymbol{\Psi}_l \star \mathbf{G}_l, \quad l = 0, \dots, L.$$

---

[5]Since A is the operator block matrix $\begin{pmatrix} \mathbf{0} & -\operatorname{div} & \cdots & -\operatorname{div} \\ -\varepsilon & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ -\varepsilon & \mathbf{0} & \cdots & \mathbf{0}\end{pmatrix}$ we see that the unique extension A_- must have zeros at the same blocks, that is, A and A_- share the same block structure.

Proof. By (13) and the definition of Φ,

$$\left[\Phi'(w,\mathbf{p})^* \begin{pmatrix} g \\ \Sigma \end{pmatrix}\right] \begin{pmatrix} \widehat{w} \\ \widehat{\mathbf{p}} \end{pmatrix} = \big\langle \widehat{w}, (I+J^*Q^*)V(\mathbf{p})^* g + A_-^* J^* g + \Psi^*\Sigma \big\rangle_{\mathcal{H}} \\ + \big\langle g, [V'(\mathbf{p})\widehat{\mathbf{p}}]\big((I+QJ)w - u_0\big)\big\rangle_{\mathcal{H}}$$

where $u_0 = (\mathbf{v}_0, \boldsymbol{\sigma}_{0,0}, \ldots, \boldsymbol{\sigma}_{L,0})^\top$ is the initial value of (4).

In view of (5), (15), and (16) we have $Q^* = Q$ as well as $V(\mathbf{p})^* = V(\mathbf{p})$. Further, $J^*QV(\mathbf{p}) = QV(\mathbf{p})J^*$ and $A_-^* = -A_-$ (recall from Lemma 4.1 that A is skew-symmetric). Hence, the stated representation of

$$\Phi'(w,\mathbf{p})_1^* \begin{pmatrix} g \\ \Sigma \end{pmatrix} = (I+Q)V(\mathbf{p})g^{\downarrow} - A_- J g^{\downarrow} + \Psi^*\Sigma$$

follows. To obtain the second component of $\Phi'(w,\mathbf{p})^*$ we evaluate

$$\big\langle g, [V'(\mathbf{p})\widehat{\mathbf{p}}]\big((I+QJ)w - u_0\big)\big\rangle_{\mathcal{H}} = \int_0^T \big\langle g(t), [V'(\mathbf{p})\widehat{\mathbf{p}}]\big(w(t) - u_0 + Qw^{\uparrow}(t)\big)\big\rangle_X \mathrm{d}t \tag{23}$$

using (19), (18), and the shortcuts introduced in (22):

$$\big\langle g(t), [V'(\mathbf{p})\widehat{\mathbf{p}}]\big(w(t) - u_0 + Qw^{\uparrow}(t)\big)\big\rangle_X = \int_D S(t,x)\mathrm{d}x \tag{24}$$

with $\widehat{\mathbf{p}} = (\widehat{\rho}, \widehat{v}_{\mathrm{S}}, \widehat{\tau}_{\mathrm{S}}, \widehat{v}_{\mathrm{P}}, \widehat{\tau}_{\mathrm{P}}) \in L^\infty(D)^5$ and

$$S(t,x) = \widehat{\rho}\,(\mathbf{w} - \mathbf{v}_0)\cdot \mathbf{g}_{-1} - \frac{\widehat{\rho}}{\rho}\boldsymbol{\Psi}_0 : \mathbf{g}_0 - \rho\, C(\widetilde{\mu}, \widetilde{\pi})\boldsymbol{\Psi}_0 : \mathbf{G}_0 \\ - \sum_{l=1}^{L} \Big(\frac{\widehat{\rho}}{\rho}\boldsymbol{\Psi}_l : \mathbf{g}_l - \rho\, C(\widehat{\mu}, \widehat{\pi})\boldsymbol{\Psi}_l : \mathbf{G}_l\Big)$$

where we suppressed the time and space dependence of the terms in boldface. Next we calculate

$$\begin{aligned} C(\widetilde{\mu}, \widetilde{\pi})\boldsymbol{\Psi}_0 : \mathbf{G}_0 &= C(a_{\mathrm{S}}\,\widehat{v}_{\mathrm{S}} - \alpha\, b_{\mathrm{S}}\,\widehat{\tau}_{\mathrm{S}}, a_{\mathrm{P}}\,\widehat{v}_{\mathrm{P}} - \alpha\, b_{\mathrm{P}}\,\widehat{\tau}_{\mathrm{P}})\boldsymbol{\Psi}_0 : \mathbf{G}_0 \\ &= 2(a_{\mathrm{S}}\,\widehat{v}_{\mathrm{S}} - \alpha\, b_{\mathrm{S}}\,\widehat{\tau}_{\mathrm{S}})\boldsymbol{\Psi}_0 : \mathbf{G}_0 \\ &\qquad + \big(a_{\mathrm{P}}\,\widehat{v}_{\mathrm{P}} - \alpha\, b_{\mathrm{P}}\,\widehat{\tau}_{\mathrm{P}} - 2(a_{\mathrm{S}}\,\widehat{v}_{\mathrm{S}} - \alpha\, b_{\mathrm{S}}\,\widehat{\tau}_{\mathrm{S}})\big)\,\mathrm{tr}(\boldsymbol{\Psi}_0)\,\mathrm{tr}(\mathbf{G}_0) \\ &= \widehat{v}_{\mathrm{S}}\, 2a_{\mathrm{S}}\,\boldsymbol{\Psi}_0 \vartriangle \mathbf{G}_0 - \widehat{\tau}_{\mathrm{S}}\, 2\alpha\, b_{\mathrm{S}}\,\boldsymbol{\Psi}_0 \vartriangle \mathbf{G}_0 \\ &\qquad + \widehat{v}_{\mathrm{P}}\, a_{\mathrm{P}}\,\boldsymbol{\Psi}_0 \star \mathbf{G}_0 - \widehat{\tau}_{\mathrm{P}}\,\alpha\, b_{\mathrm{P}}\,\boldsymbol{\Psi}_0 \star \mathbf{G}_0 \end{aligned}$$

and, similarly,

$$C(\widehat{\mu}, \widehat{\pi})\boldsymbol{\Psi}_l : \mathbf{G}_l = \widehat{v}_{\mathrm{S}}\, 2\tau_{\mathrm{S}}\, a_{\mathrm{S}}\,\boldsymbol{\Psi}_l \vartriangle \mathbf{G}_l + \widehat{\tau}_{\mathrm{S}}\, 2b_{\mathrm{S}}\,\boldsymbol{\Psi}_l \vartriangle \mathbf{G}_l \\ + \widehat{v}_{\mathrm{P}}\,\tau_{\mathrm{P}}\, a_{\mathrm{P}}\,\boldsymbol{\Psi}_l \star \mathbf{G}_l + \widehat{\tau}_{\mathrm{P}}\, b_{\mathrm{P}}\,\boldsymbol{\Psi}_l \star \mathbf{G}_l.$$

We plug both these expressions into S and sort the result by the components of $\widehat{\mathbf{p}}$:

$$S = \widehat{\rho}\Big((\mathbf{w}-\mathbf{v}_0)\cdot \mathbf{g}_{-1} - \frac{1}{\rho}\,\boldsymbol{\Psi}_0 : \mathbf{g}_0 - \frac{1}{\rho}\sum_{l=1}^{L}\boldsymbol{\Psi}_l : \mathbf{g}_l\Big) - \widehat{v}_\mathrm{S}\, 2a_\mathrm{S}\,\rho\Big(\boldsymbol{\Psi}_0 \vartriangle \mathbf{G}_0 + \tau_\mathrm{S}\sum_{l=1}^{L}\boldsymbol{\Psi}_l \vartriangle \mathbf{G}_l\Big)$$
$$+\, \widehat{\tau}_\mathrm{S}\, 2b_\mathrm{S}\,\rho\Big(\alpha\boldsymbol{\Psi}_0 \vartriangle \mathbf{G}_0 - \sum_{l=1}^{L}\boldsymbol{\Psi}_l \vartriangle \mathbf{G}_l\Big) - \widehat{v}_\mathrm{P}\, a_\mathrm{P}\,\rho\Big(\boldsymbol{\Psi}_0 \star \mathbf{G}_0 + \tau_\mathrm{P}\sum_{l=1}^{L}\boldsymbol{\Psi}_l \star \mathbf{G}_l\Big)$$
$$+\, \widehat{\tau}_\mathrm{P}\, b_\mathrm{P}\,\rho\Big(\alpha\boldsymbol{\Psi}_0 \star \mathbf{G}_0 - \sum_{l=1}^{L}\boldsymbol{\Psi}_l \star \mathbf{G}_l\Big).$$

Having this representation of S in mind, recalling (24) and changing the order of integration in (23), we finally obtain the stated form of $\Phi'(w,\mathbf{p})_2^*$. ☐

Remark 4.5. *The previous two propositions are valid also for the viscoelastic equation in two spatial dimensions. However, there is a difference in the representation of* $\widetilde{C} = C^{-1}$*. Since* $\mathrm{tr}(\mathbf{I}) = 2$ *in 2D we have that*

$$\widetilde{C}(m,p)\mathbf{M} = C\Big(\frac{1}{4m}, \frac{p}{4m(p-m)}\Big)\mathbf{M} = \frac{1}{2m}\,\mathbf{M} + \frac{2m-p}{4m(p-m)}\,\mathrm{tr}(\mathbf{M})\mathbf{I},$$

compare (14)*.*

References

[1] H. S. Aghamiry, A. Gholami, and S. Operto, *Compound regularization of full-waveform inversion for imaging piecewise media*, IEEE Trans. Geosci. Remote Sensing, 58 (2020), pp. 1192–1204, https://doi.org/10.1109/TGRS.2019.2944464.

[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, *Viskoelastische FD-Modellierung seismischer Wellen zur Interpretation gemessener Seismogramme*, PhD thesis, Christian-Albrechts-Universität zu Kiel, 1998, https://bit.ly/2LMOSWr.

[4] T. Bohlen, *Parallel 3-D viscoelastic finite difference seismic modelling*, Comput. Geosci., 28 (2002), pp. 887–899, https://doi.org/10.1016/S0098-3004(02)00006-7.

[5] T. Bohlen, M. R. Fernandez, J. Ernesti, C. Rheinbay, A. Rieder, and C. Wieners, *Visco-acoustic full waveform seismic inversion: from a DG forward solver to a Newton-CG inverse solver*, CRC 1173 Preprint 2020/4, Karlsruhe Institute of Technology, feb 2020, https://doi.org/10.5445/IR/1000105695.

[6] K.-J. Engel and R. Nagel, *One-parameter semigroups for linear evolution equations*, vol. 194 of Graduate Texts in Mathematics, Springer-Verlag, New York, 2000, https://doi.org/10.1007/b97696.

[7] A. Fichtner, *Full Seismic Waveform Modelling and Inversion*, Advances in Geophysical and Environmental Mechanics and Mathematics, Springer-Verlag Berlin Heidelberg, 2011, https://doi.org/10.1007/978-3-642-15807-0.

[8] E. Haber and U. M. Ascher, *Preconditioned all-at-once methods for large, sparse parameter estimation problems*, Inverse Problems, 17 (2001), pp. 1847–1864, https://doi.org/10.1088/0266-5611/17/6/319.

[9] E. Haber, U. M. Ascher, and D. W. Oldenburg, *Inversion of 3d electromagnetic data in frequency and time domain using an inexact all-at-once approach*, Geophysics, 69 (2004), pp. 1216–1228, https://doi.org/10.1190/1.1801938.

[10] B. Hofmann, *On ill-posedness and local ill-posedness of operator equations in Hilbert spaces*, Tech. Report 97-8, Fakultät für Mathematik, Technische Universität Chemnitz-Zwickau, D-09107 Chemnitz, 1997, http://www.qucosa.de/fileadmin/data/qucosa/documents/4197/data/a008.pdf.

[11] B. Kaltenbacher, *Regularization based on all-at-once formulations for inverse problems*, SIAM J. Numer. Anal., 54 (2016), pp. 2594–2618, https://doi.org/10.1137/16M1060984.

[12] B. Kaltenbacher, *All-at-once versus reduced iterative methods for time dependent inverse problems*, Inverse Problems, 33 (2017), pp. 064002, 31, https://doi.org/10.1088/1361-6420/aa6f34.

[13] 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.

[14] 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.

[15] R. Schnaubelt, *Evolution equations*, lecture notes, Karlsruhe Institute of Technology, 2019, http://www.math.kit.edu/iana3/~schnaubelt/media/nevgl-skript.pdf.

[16] A. Shenoy, M. Heinkenschloss, and E. M. Cliff, *Airfoil design by an all-at-once method*, Int. J. Comp. Fluid Dynamics, 11 (1998), pp. 3–25, https://doi.org/10.1080/10618569808940863.

[17] T. van Leeuwen and F. J. Herrmann, *Mitigating local minima in full-waveform inversion by expanding the search space*, Geophys. J. Int., 195 (2013), pp. 661–667, https://doi.org/10.1093/gji/ggt258.

[18] T. van Leeuwen and F. J. Herrmann, *A penalty method for PDE-constrained optimization in inverse problems*, Inverse Problems, 32 (2016), pp. 015007, 26, https://doi.org/10.1088/0266-5611/32/1/015007.

[19] J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics*, Geophysics, 74 (2009), pp. 1–26, https://doi.org/10.1190/1.3238367.

[20] 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.

Department of Mathematics, Karlsruhe Institute of Technology (KIT), D-76128 Karlsruhe, Germany

E-mail address: andreas.rieder@kit.edu