Received: 25 May 2020 | Revised: 25 September 2020 | Accepted: 24 December 2020

DOI: 10.1002/mma.7190

RESEARCH ARTICLE

WILEY

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

Andreas Rieder

Department of Mathematics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Correspondence
Andreas Rieder, Department of Mathematics, Karlsruhe Institute of Technology (KIT), Karlsruhe D-76128, Germany.
Email: andreas.rieder@kit.edu

Communicated by: B. Harrach

Funding information
Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Grant/Award Number: Project-ID 258734477 - SFB 1173

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.

KEYWORDS

adjoint state method, all-at-once method, full waveform inversion, viscoelastic wave equation, wavefield reconstruction inversion

MSC CLASSIFICATION

35F10; 35R30; 86A22

1 | INTRODUCTION

Parameter identification problems for partial differential equations (pdes) 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, for example, previous studies.[1-3] 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 Kaltenbacher,[4] 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.[5]

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

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

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

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

In this work, we give a mathematically clean all-at-once formulation for time-domain full waveform inversion (FWI) in the viscoelastic regime. Moreover, we provide necessary ingredients to set up Newton/gradient-type regularization schemes. These ingredients are the Fréchet derivative $\Phi'_{\text{all}}(u, p)$ of the corresponding nonlinear map Φ_{all} as well as an explicit representation of the adjoint operator $\Phi'_{\text{all}}(u, p)^*$.

FWI is the state-of-the-art inversion procedure in geophysical exploration taking full advantage of the amplitude and phase information of seismic recordings; see, for example, Fichtner and Virieux and Operto.[6,7] 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 van Leeuwen and Herrmann,[8] however, for the acoustic wave equation in the frequency domain (Helmholtz equation). In the subsequent publication,[9] WRI has been generalized to other nondynamic (frequency domain) inverse problems. We refer to Aghamiry et al.[10] 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 which is in contrast to our infinite-dimensional setting.

The advantages of the all-at-once approach to FWI reported and demonstrated in the above cited literature are as follows:

- 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 (Lemma 3.5), and provide a representation of the adjoint operator of the Fréchet derivative (Lemma 3.6). 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 (Propositions 4.3 and 4.4). 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, for example, Kirsch and Rieder.[11] We need to recall it nevertheless for sake of completeness, for introducing some notation, and for indicating a minor flaw; see Remark 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 Fichtner[6], Chap. 5 and Zeltmann[12], 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 $\boldsymbol{\eta}_l : [0, T] \times D \to \mathbb{R}^{3\times 3}_{\text{sym}}$, $l = 1, \dots, L$, the new formulation reads

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

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

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

where $D \subset \mathbb{R}^3$ is a Lipschitz domain. The functions $\tau_P, \tau_{\mathrm{S}} : D \to \mathbb{R}$ are scaling factors for the unrelaxed bulk moduls π and shear modulus μ, respectively. They have been introduced by Blanch et al.[13]

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

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

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 $\mathrm{tr}(\mathbf{M})$ denotes the trace of $\mathbf{M} \in \mathbb{R}^{3\times 3}$. Finally,

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

is the linearized strain rate.

Wave propagation is frequency-dependent, and the numbers $\tau_{\sigma,l} > 0$, $l = 1,\dots,L$, 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 $\tau_{\sigma,l}$ by a least-squares approach.[14,15]

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

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

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

By the transformation, see Zeltmann,[12]

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

we reformulate (1) equivalently into

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

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

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

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* X and suitable $w = (\mathbf{w}, \psi_0, \dots, \psi_L) \in X$, we define operators A, B, and Q mapping into X by

$$Aw = -\begin{pmatrix} \mathrm{div}\left(\sum_{l=0}^{L}\psi_l\right) \\ \varepsilon(\mathbf{w}) \\ \vdots \\ \varepsilon(\mathbf{w}) \end{pmatrix},\ B^{-1}w = \begin{pmatrix} \frac{1}{\rho}\mathbf{w} \\ C(\mu,\pi)\psi_0 \\ C(\tau_{\mathrm{S}}\mu,\tau_P\pi)\psi_1 \\ \vdots \\ C(\tau_{\mathrm{S}}\mu,\tau_P\pi)\psi_L \end{pmatrix},\ Qw = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \frac{1}{\tau_{\sigma,1}}\psi_1 \\ \vdots \\ \frac{1}{\tau_{\sigma,L}}\psi_L \end{pmatrix}. \tag{5}$$

*Details follow below.

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 $f = (f, \mathbf{0}, \dots, \mathbf{0})$.

The operator B depends solely on the five parameters $(\rho, v_S, \tau_S, v_P, \tau_P)$ since, according to (3),

$$\pi = \frac{\rho v_P^2}{1+\tau_P\alpha} \quad \text{and} \quad \mu = \frac{\rho v_S^2}{1+\tau_S\alpha}. \tag{6}$$

In a first step, we "hide" these parameters in the operator B which we consider—for the time being—as the searched-for object (Section 3). Finally, in Section 4.2, we include the mapping $(\rho, v_S, \tau_S, v_P, \tau_P) \mapsto B$ into our analysis of the viscoelastic system (4).

Remark 1. Please note that in the previous publication,[11] 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

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

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

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

In Kirsch and Rieder,[11] 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, for example, Bohlen et al.[16] The viscoacoustic equation models wave propagation in media which do not support shear stress.[†]

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 Kirsch and Rieder.[11,17] Hence, standard existence and uniqueness results apply, for instance, (7) has a unique integrated solution $u \in C([0, T], X)$ satisfying $\int_0^t u(s)\,\mathrm{d}s \in \mathsf{D}(A)$, $t \in]0, T[$, and

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

which coincides with the mild solution; see, for example, Schnaubelt.[18], Prop. 2.15

Next we extend A by $A_- : \mathsf{D}(A_-) \subset X_- \to X_-$ to X_- which is the completion[‡] 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, for example, Engel and Nagel[19], Chap. 2.5 and Schnaubelt[18], Chap. 2.2 for the details.

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

[‡]The standard notation for X_- is X_{-1} (as an element of a Sobolev tower/scale[19], Chap. 2.5).

Using A_-, we generalize (8) slightly to

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

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

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

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

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

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

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

The terms on the right hand side converge in X_- when $h \to 0$. Indeed, $1/h\int_t^{t+h} f(s)\,\mathrm{d}s$ and $BQ1/h\int_t^{t+h} u(s)\,\mathrm{d}s$ 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 $A_-u(t)$. □

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 does not lie 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 $A : \mathsf{D}(A) \subset X \to X$.

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 $\Psi : \mathcal{H} \to \mathbb{R}^{\boldsymbol{N}}$, $\boldsymbol{N} \in \mathbb{N}$, 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: $Jv(t) = \int_0^t v(s)\mathrm{d}s$.

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

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

where

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

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 $\Sigma \in \mathbb{R}^{\boldsymbol{N}}$, then $u \in C([0,T],X)$ 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 $\Sigma \in \mathbb{R}^{\boldsymbol{N}}$ (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 $F_{\text{red}} : B \mapsto \Psi u$ 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

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

Please observe that $F_1 : \mathcal{H} \times \mathcal{L}^*(X) \to \mathcal{H} \times \mathbb{R}^{\boldsymbol{N}}$ is bilinear and bounded, $F_2 \in \mathcal{L}(\mathcal{H} \times \mathcal{L}^*(X), \mathcal{H}_- \times \mathbb{R}^{\boldsymbol{N}})$, 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*

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

The second derivative is given by

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

All higher derivatives vanish identically.

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

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

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

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

Proof. We have

$$\begin{aligned} \left[F'(v,P)^*\begin{pmatrix} g \\ \Sigma \end{pmatrix}\right]\begin{pmatrix} \hat{v} \\ \hat{P} \end{pmatrix} &= \left\langle \begin{pmatrix} g \\ \Sigma \end{pmatrix}, F'(v,P)\begin{pmatrix} \hat{v} \\ \hat{P} \end{pmatrix}\right\rangle_{(\mathcal{H}'_- \times \mathbb{R}^{\boldsymbol{N}}) \times (\mathcal{H}_- \times \mathbb{R}^{\boldsymbol{N}})} \\ &= \langle g, P(I+QJ)\hat{v} + A\_J\hat{v}\rangle_{\mathcal{H}} + \left\langle g, \hat{P}(I+QJ)v - \hat{P}u_0\right\rangle_{\mathcal{H}} + \langle \Sigma, \Psi\hat{v}\rangle_{\mathbb{R}^{\boldsymbol{N}}} \end{aligned} \tag{13}$$

from which the assertion follows immediately. □

4 | APPLICATION TO FWI

4.1 | The setting and basic definitions

We recall the basic concepts from Kirsch and Rieder.[11]

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

$$\langle(\mathbf{v},\boldsymbol{\sigma}_0,\dots,\boldsymbol{\sigma}_L),(\mathbf{w},\psi_0,\dots,\psi_l)\rangle_X = \int_D \left(\mathbf{v}\cdot\mathbf{w} + \sum_{l=0}^{L}\boldsymbol{\sigma}_l : \psi_l\right)\mathrm{d}x.$$

The colon represents the Frobenius inner product of matrices.

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

$$\mathsf{D}(A) = \left\{ (\mathbf{w}, \psi_0, \ldots \psi_L) \in H_D^1 \times H(\mathrm{div})^{1+L} : \sum_{l=0}^{L} \psi_l \boldsymbol{n} = 0 \text{ on } \partial D_{\boldsymbol{N}} \right\}$$

for the domain of A from (5). Here, $H_D^1 = \{\mathbf{v} \in H^1(D, \mathbb{R}^3) : \mathbf{v} = 0 \text{ on } \partial D_D\}$ and $H(\mathrm{div}) = \left\{ \boldsymbol{\sigma} \in L^2\left(D, \mathbb{R}^{3\times 3}_{\mathrm{sym}}\right) : \mathrm{div}\left(\sum_{l=0}^{L} \boldsymbol{\sigma}_l\right) \in L^2(D) \right\}$. The following result is validated in Kirsch A, Rieder.[11], 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, that is, $A^* = -A$.*

Next we present the representation of B for the viscoelastic setting. A crucial ingredient is the operator C of (2) which maps $\mathsf{D}(C) = \left\{ (m, p) \in \mathbb{R}^2 : \underline{\mathrm{m}} \le m \le \overline{\mathrm{m}}, \underline{\mathrm{p}} \le p \le \overline{\mathrm{p}} \right\}$ into $\mathrm{Aut}(\mathbb{R}^{3\times 3}_{\mathrm{sym}})$[§] 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 $(m, p) \in \mathsf{D}(C)$,

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

We have

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

and

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

see, for example, Zeltmann.[12], Lemma 50. If $\rho(x) > 0$, $(\mu(x), \pi(x))$, $(\tau_S(x)\mu(x), \tau_P(x)\pi(x)) \in \mathsf{D}(C)$ for almost all $x \in D$, then

$$B \begin{pmatrix} \mathbf{w} \\ \psi_0 \\ \psi_1 \\ \vdots \\ \psi_L \end{pmatrix} = \begin{pmatrix} \rho \mathbf{w} \\ \tilde{C}(\mu, \pi)\psi_0 \\ \tilde{C}(\tau_S \mu, \tau_P \pi)\psi_1 \\ \vdots \\ \tilde{C}(\tau_S \mu, \tau_P \pi)\psi_L \end{pmatrix} \tag{16}$$

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 $\mathbf{p} = (\rho, v_S, \tau_S, v_P, \tau_P)$ from observed wavefields. Therefore, we here define an all-at-once operator $\Phi(w, \mathbf{p}) := F(w, V(\mathbf{p}))$ where $V : \mathbf{p} \mapsto B$ and F is the mapping from (11).

A physically meaningful domain of definition for V is

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

with suitable positive bounds $0 < \rho_{\min} < \rho_{\max} < \infty$, and so forth. Note that $\underline{\mathrm{p}}$, $\overline{\mathrm{p}}$, $\underline{\mathrm{m}}$, and $\overline{\mathrm{m}}$ can be defined in terms of $\rho_{\min}, \rho_{\max}, v_{P,\min}$, and so forth such that $(\mu, \pi), (\tau_S \mu, \tau_P \pi)$ as functions of $(\rho, v_P, v_S, \tau_P, \tau_S) \in \mathsf{D}(V)$ are in $\mathsf{D}(C)$; see Kirsch and Rieder.[11]

Hence, we have a well-defined mapping

$$V : \mathsf{D}(V) \subset L^\infty(D)^5 \to \mathcal{B} \subset \mathcal{L}^*(X), \ (\rho, v_S, \tau_S, v_P, \tau_P) \mapsto B,$$

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

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

$$\Phi : \mathsf{D}(\Phi) \subset \mathcal{H} \times L^\infty(D)^5 \to \mathcal{H}_- \times \mathbb{R}^{\boldsymbol{N}}, \tag{17a}$$

$$(w, (\rho, \nu_S, \tau_S, \nu_P, \tau_P)) \mapsto F\left(w, V(\rho, \nu_S, \tau_S, \nu_P, \tau_P)\right), \tag{17b}$$

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

$$\Phi(u, \mathbf{p}) = \begin{pmatrix} 0 \\ \Sigma \end{pmatrix}. \tag{18}$$

If $(u, \mathbf{p})^\top$ solves the above equation, then $u = (\mathbf{v}, \boldsymbol{\sigma}_0, \dots, \boldsymbol{\sigma}_L)^\top$ 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 Kirsch and Rieder,[11], Theorem 4.3, the all-at-once version (18) is locally ill-posed. Local ill-posedness of an inverse problem was introduced in Hofmann:[20] Let $\Theta : \mathsf{D}(\Theta) \subset V \to W$ 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 $D(\Phi)$.*

Proof. We adapt the proof of Theorem 4.3 of Kirsch and Rieder[11] to the present situation. Choose an $x_o \in D$ and define the balls $K_k = \{x \in \mathbb{R}^3 : |x - x_o| \le \epsilon/k\}$, $k \in \mathbb{N}$, 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, $\{(u_k, \mathbf{p}_k)\} \subset \mathsf{D}(\Phi)$, $(u, \mathbf{p}) \in \mathsf{D}(\Phi)$, $\Phi_1(u_k, \mathbf{p}_k) = 0 = \Phi_1(u, \mathbf{p})$, 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 a related evolution equation and then by applying the stability estimate for mild solutions, see, for example, Kirsch and Rieder[11] 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 (18) 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 Kirsch and Rieder.[11] The formulation of V' relies on the derivative of $\tilde{C}$ which is

$$\tilde{C}'(m,p)\begin{bmatrix} \hat{m} \\ \hat{p} \end{bmatrix} = -\tilde{C}(m,p) \circ C(\hat{m}, \hat{p}) \circ \tilde{C}(m,p) \tag{19}$$

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

Let $\mathbf{p} = (\rho, \nu_S, \tau_S, \nu_P, \tau_P) \in \mathrm{int}(\mathsf{D}(\Phi))$ and $\hat{\mathbf{p}} = (\hat\rho, \hat\nu_S, \hat\tau_S, \hat\nu_P, \hat\tau_P) \in L^\infty(D)^5$. Then, $V'(\mathbf{p})\hat{\mathbf{p}} \in \mathcal{L}^*(X)$ is given by

$$V'(\mathbf{p})\hat{\mathbf{p}} \begin{pmatrix} \mathbf{w} \\ \psi_0 \\ \vdots \\ \psi_L \end{pmatrix} = \begin{pmatrix} \hat\rho \mathbf{w} \\ -\frac{\hat\rho}{\rho}\tilde C(\mu,\pi)\psi_0 + \rho \tilde C'(\mu,\pi)\begin{bmatrix}\tilde\mu \\ \tilde\pi\end{bmatrix}\psi_0 \\ -\frac{\hat\rho}{\rho}\tilde C(\tau_S\mu,\tau_P\pi)\psi_1 + \rho \tilde C'(\tau_S\mu,\tau_P\pi)\begin{bmatrix}\hat\mu \\ \hat\pi\end{bmatrix}\psi_1 \\ \vdots -\frac{\hat\rho}{\rho}\tilde C(\tau_S\mu,\tau_P\pi)\psi_L + \rho \tilde C'(\tau_S\mu,\tau_P\pi)\begin{bmatrix}\hat\mu \\ \hat\pi\end{bmatrix}\psi_L \end{pmatrix} \tag{20}$$

where μ and π are from (6) and

$$\tilde\mu = a_S\hat\nu_S - \alpha b_S\hat\tau_S, \quad \tilde\pi = a_P\hat\nu_P - \alpha b_P\hat\tau_P, \tag{21}$$

$$\hat\mu = \tau_S a_S\hat\nu_S + b_S\hat\tau_S, \quad \hat\pi = \tau_P a_P\hat\nu_P + b_P\hat\tau_P, \tag{22}$$

with

$$a_S = \frac{2\nu_S}{1+\tau_S\alpha}, \; b_S = \frac{\nu_S^2}{(1+\tau_S\alpha)^2}, \; a_P = \frac{2\nu_P}{1+\tau_P\alpha}, \; b_P = \frac{\nu_P^2}{(1+\tau_P\alpha)^2}.$$

We introduce new symbolic notation:

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

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 $\mathsf{D}(\Phi)$, $w = (\mathbf{w}, \psi_0, \ldots, \psi_L)$, $\mathbf{p} = (\rho, \nu_S, \tau_S, \nu_P, \tau_P)$:*
For $\hat w = (\hat{\mathbf{w}}, \hat\psi_0, \ldots, \hat\psi_L) \in \mathcal{H}$ and $\hat{\mathbf{p}} = (\hat\rho, \hat\nu_S, \hat\tau_S, \hat\nu_P, \hat\tau_P) \in L^\infty(D)^5$, we have

$$\Phi'(w,\mathbf{p})\begin{bmatrix}\hat w \\ \hat{\mathbf{p}}\end{bmatrix} = \begin{pmatrix} \rho\hat{\mathbf{w}} + \hat\rho(\mathbf{w}-\mathbf{v}_0) - div_-\left(\sum_{l=0}^L \hat\psi_l^\uparrow\right) \\ \tilde C(\mu,\pi)\left(\hat\psi_0 - \frac{\hat\rho}{\rho}(\psi_0 - \boldsymbol{\sigma}_{0,0})\right) + \rho\tilde C'(\mu,\pi)\begin{bmatrix}\tilde\mu \\ \tilde\pi\end{bmatrix}(\psi_0 - \boldsymbol{\sigma}_{0,0}) - \varepsilon_-(\hat{\mathbf{w}}^\uparrow) \\ \tilde C(\tau_S\mu,\tau_P\pi)\left(\left(1+\frac{1}{\tau_{\sigma,1}}\right)\hat\psi_1^\uparrow - \frac{\hat\rho}{\rho}\left(\psi_1 - \boldsymbol{\sigma}_{1,0} + \frac{1}{\tau_{\sigma,1}}\psi_1^\uparrow\right)\right) \\ +\rho\tilde C'(\tau_S\mu,\tau_P\pi)\begin{bmatrix}\hat\mu \\ \hat\pi\end{bmatrix}\left(\psi_1 - \boldsymbol{\sigma}_{1,0} + \frac{1}{\tau_{\sigma,1}}\psi_1^\uparrow\right) - \varepsilon_-(\hat{\mathbf{w}}^\uparrow) \\ \vdots \\ \tilde C(\tau_S\mu,\tau_P\pi)\left(\left(1+\frac{1}{\tau_{\sigma,L}}\right)\hat\psi_L^\uparrow - \frac{\hat\rho}{\rho}\left(\psi_L - \boldsymbol{\sigma}_{L,0} + \frac{1}{\tau_{\sigma,L}}\psi_L^\uparrow\right)\right) \\ +\rho\tilde C'(\tau_S\mu,\tau_P\pi)\begin{bmatrix}\hat\mu \\ \hat\pi\end{bmatrix}\left(\psi_L - \boldsymbol{\sigma}_{L,0} + \frac{1}{\tau_{\sigma,L}}\psi_L^\uparrow\right) - \varepsilon_-(\hat{\mathbf{w}}^\uparrow) \\ \Psi(\hat{\mathbf{w}}, \hat\psi_0, \ldots, \hat\psi_L) \end{pmatrix}$$

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

¶ Since A is the operator block matrix $\begin{pmatrix} \mathbf{0} & -\mathrm{div} & \ldots & -\mathrm{div} \\ -\varepsilon & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ -\varepsilon & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix}$, we see that the unique extension A_- must have zeros at the same blocks; that is, A and A_- share the same block structure. Moreover, as the definitions of the differential operators div and ε are independent of their block positions in A, only their extensions div_- and ε_- are needed to make up A_-.

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

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

Finally, we plug in the expressions from (5), (16), and (20). □

Proposition 4.4. *The notation and the assumptions are as in the previous proposition. Then, the adjoint* $\Phi'(w,\mathbf{p})^* \in \mathcal{L}\left(\mathcal{H}'_{-}\times\mathbb{R}^{\mathbf{N}}, \mathcal{H}'\times(L^\infty(D)^5)'\right)$ *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 $g = (g_{-1}, g_0, \dots, g_L) \in L^2\left([0,T], L^2(D,\mathbb{R}^3)\times L^2(D,\mathbb{R}^{3\times3}_{sym})^{1+L}\right)$, $\Sigma\in\mathbb{R}^{\mathbf{N}}$,

$$\Phi'(w,\mathbf{p})_1^*\begin{pmatrix} g\\ \Sigma\end{pmatrix} = \begin{pmatrix} \rho g_{-1} + div_{-}\left(\sum_{l=0}^{L} g_l^{\downarrow}\right) \\ \tilde{C}(\mu,\pi)g_0 + \varepsilon_{-}(g_{-1}^{\downarrow}) \\ \tilde{C}(\tau_S\mu,\tau_P\pi)\left(g_1 + \frac{1}{\tau_{\sigma,1}}g_1^{\downarrow}\right) + \varepsilon_{-}(g_{-1}^{\downarrow}) \\ \vdots \\ \tilde{C}(\tau_S\mu,\tau_P\pi)\left(g_L + \frac{1}{\tau_{\sigma,L}}g_L^{\downarrow}\right) + \varepsilon_{-}(g_{-1}^{\downarrow}) \end{pmatrix}^{\top} + \Psi^*\Sigma$$

and

$$\Phi'(w,\mathbf{p})_2^*\begin{pmatrix} g\\ \Sigma\end{pmatrix} = \begin{pmatrix} \int_0^T\left((\mathbf{w}-\mathbf{v}_0)\cdot g_{-1} - \frac{1}{\rho}\mathbf{\Psi}_0 : g_0 - \frac{1}{\rho}\sum_{l=1}^{L}\mathbf{\Psi}_l : g_l\right)\mathrm{d}t \\ -2a_S\,\rho\int_0^T\left(\mathbf{\Psi}_0 \vartriangle G_0 + \tau_S\sum_{l=1}^{L}\mathbf{\Psi}_l \vartriangle G_l\right)\mathrm{d}t \\ 2b_S\,\rho\int_0^T\left(\alpha\mathbf{\Psi}_0 \vartriangle G_0 - \sum_{l=1}^{L}\mathbf{\Psi}_l \vartriangle G_l\right)\mathrm{d}t \\ -a_P\,\rho\int_0^T\left(\mathbf{\Psi}_0 \star G_0 + \tau_P\sum_{l=1}^{L}\mathbf{\Psi}_l \star G_l\right)\mathrm{d}t \\ b_P\,\rho\int_0^T\left(\alpha\mathbf{\Psi}_0 \star G_0 - \sum_{l=1}^{L}\mathbf{\Psi}_l \star G_l\right)\mathrm{d}t \end{pmatrix}^{\top}$$

with the following abbreviations

$$\mathbf{\Psi}_0 = \tilde{C}(\mu,\pi)(\psi_0 - \boldsymbol{\sigma}_{0,0}), \qquad \mathbf{\Psi}_l = \tilde{C}(\tau_S\mu,\tau_P\pi)(\psi_l - \boldsymbol{\sigma}_{l,0} + \psi_l^{\uparrow}/\tau_{\sigma,l}),\ l = 1,\dots,L, \tag{23a}$$

$$G_0 = \tilde{C}(\mu,\pi)g_0, \qquad G_l = \tilde{C}(\tau_S\mu,\tau_P\pi)g_l,\ l=1,\dots,L, \tag{23b}$$

and

$$\mathbf{\Psi}_l \star G_l = tr\,(\mathbf{\Psi}_l)\ tr\,(G_l),\ \ \mathbf{\Psi}_l \vartriangle G_l = \mathbf{\Psi}_l : G_l - \mathbf{\Psi}_l \star G_l,\ \ l = 0,\dots,L.$$

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

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

where $u_0 = (\mathbf{v}_0, \boldsymbol{\sigma}_{0,0}, \dots, \boldsymbol{\sigma}_{L,0})^\top$ 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, $J^* Q V(\mathbf{p}) = Q V(\mathbf{p}) J^*$ 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

$$\left\langle g, [V'(\mathbf{p})\widehat{\mathbf{p}}]\left((I + QJ)w - u_0\right)\right\rangle_{\mathcal{H}} = \int_0^T \left\langle g(t), [V'(\mathbf{p})\widehat{\mathbf{p}}]\left(w(t) - u_0 + Q w^{\uparrow}(t)\right)\right\rangle_X \mathrm{d}t \tag{24}$$

using (20), (19), and the shorthand notations introduced in (23):

$$\left\langle g(t), [V'(\mathbf{p})\widehat{\mathbf{p}}]\left(w(t) - u_0 + Q w^{\uparrow}(t)\right)\right\rangle_X = \int_D S(t, x)\mathrm{d}x \tag{25}$$

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

$$\begin{aligned} S(t, x) = \hat{\rho}(\mathbf{w} - \mathbf{v}_0) \cdot g_{-1} - \frac{\hat{\rho}}{\rho} \boldsymbol{\Psi}_0 : g_0 - \rho C(\tilde{\mu}, \tilde{\pi}) \boldsymbol{\Psi}_0 : G_0 \\ - \sum_{l=1}^{L} \left(\frac{\hat{\rho}}{\rho} \boldsymbol{\Psi}_l : g_l - \rho C(\hat{\mu}, \hat{\pi}) \boldsymbol{\Psi}_l : G_l \right) \end{aligned}$$

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

$$\begin{aligned} C(\tilde{\mu}, \tilde{\pi}) \boldsymbol{\Psi}_0 : G_0 &= C(a_\mathrm{S} \hat{v}_\mathrm{S} - \alpha b_\mathrm{S} \hat{\tau}_\mathrm{S}, a_\mathrm{P} \hat{v}_\mathrm{P} - \alpha b_\mathrm{P} \hat{\tau}_\mathrm{P}) \boldsymbol{\Psi}_0 : G_0 \\ &= 2(a_\mathrm{S} \hat{v}_\mathrm{S} - \alpha b_\mathrm{S} \hat{\tau}_\mathrm{S}) \boldsymbol{\Psi}_0 : G_0 \\ &\qquad + (a_\mathrm{P} \hat{v}_\mathrm{P} - \alpha b_\mathrm{P} \hat{\tau}_\mathrm{P} - 2(a_\mathrm{S} \hat{v}_\mathrm{S} - \alpha b_\mathrm{S} \hat{\tau}_\mathrm{S})) \operatorname{tr}(\boldsymbol{\Psi}_0) \operatorname{tr}(G_0) \\ &= \hat{v}_\mathrm{S} 2 a_\mathrm{S} \boldsymbol{\Psi}_0 \vartriangle G_0 - \hat{\tau}_\mathrm{S} 2 \alpha b_\mathrm{S} \boldsymbol{\Psi}_0 \vartriangle G_0 \\ &\qquad + \hat{v}_\mathrm{P} a_\mathrm{P} \boldsymbol{\Psi}_0 \star G_0 - \hat{\tau}_\mathrm{P} \alpha b_\mathrm{P} \boldsymbol{\Psi}_0 \star G_0 \end{aligned}$$

and, similarly,

$$\begin{aligned} C(\hat{\mu}, \hat{\pi}) \boldsymbol{\Psi}_l : G_l = \hat{v}_\mathrm{S} 2 \tau_\mathrm{S} a_\mathrm{S} \boldsymbol{\Psi}_l \vartriangle G_l + \hat{\tau}_\mathrm{S} 2 b_\mathrm{S} \boldsymbol{\Psi}_l \vartriangle G_l \\ + \hat{v}_\mathrm{P} \tau_\mathrm{P} a_\mathrm{P} \boldsymbol{\Psi}_l \star G_l + \hat{\tau}_\mathrm{P} b_\mathrm{P} \boldsymbol{\Psi}_l \star G_l. \end{aligned}$$

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

$$\begin{aligned} S = \hat{\rho} \left((\mathbf{w} - \mathbf{v}_0) \cdot g_{-1} - \frac{1}{\rho} \boldsymbol{\Psi}_0 : g_0 - \frac{1}{\rho} \sum_{l=1}^{L} \boldsymbol{\Psi}_l : g_l \right) - \hat{v}_\mathrm{S} 2 a_\mathrm{S} \rho \left(\boldsymbol{\Psi}_0 \vartriangle G_0 + \tau_\mathrm{S} \sum_{l=1}^{L} \boldsymbol{\Psi}_l \vartriangle G_l \right) \\ + \hat{\tau}_\mathrm{S} 2 b_\mathrm{S} \rho \left(\alpha \boldsymbol{\Psi}_0 \vartriangle G_0 - \sum_{l=1}^{L} \boldsymbol{\Psi}_l \vartriangle G_l \right) - \hat{v}_\mathrm{P} a_\mathrm{P} \rho \left(\boldsymbol{\Psi}_0 \star G_0 + \tau_\mathrm{P} \sum_{l=1}^{L} \boldsymbol{\Psi}_l \star G_l \right) \\ + \hat{\tau}_\mathrm{P} b_\mathrm{P} \rho \left(\alpha \boldsymbol{\Psi}_0 \star G_0 - \sum_{l=1}^{L} \boldsymbol{\Psi}_l \star G_l \right). \end{aligned}$$

Having this representation of S in mind, recalling (25) and changing the order of integration in (24), 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 $\tilde{C} = C^{-1}$. Since $\mathrm{tr}(\boldsymbol{I}) = 2$ in 2D, we have that

$$\tilde{C}(m,p)\mathbf{M} = C\left(\frac{1}{4m}, \frac{p}{4m(p-m)}\right)\mathbf{M} = \frac{1}{2m}\mathbf{M} + \frac{2m-p}{4m(p-m)}\,\mathrm{tr}\,(\boldsymbol{M})\boldsymbol{I},$$

compare (14).

5 | CONCLUDING REMARKS

In this work, we presented an all-at-once formulation of FWI under the viscoelastic regime in the time domain. Mathematically, this entails the inverse problem of reconstructing five material parameter functions of the viscoelastic wave equation (4) from partial measurements of reflected wavefields. We defined a corresponding forward operator Φ (17) and showed its well-posedness within a functional analytic framework. Further, we provided explicit representations of the Fréchet derivative Φ' and its adjoint operator Φ'^*. Now, these representations can be employed in a Newton/steepest decent-like scheme to numerically solve the inverse problem (18) of FWI. In contrast to the traditional, reduced FWI formulation, the all-at-once approach does not require a numerical solution of the wave and adjoint wave equations per iteration step. However, future numerical experiments must show how the new approach performs in practice. The analytic foundation for those experiments has been established in the previous sections.

ACKNOWLEDGEMENT

The author thanks Andreas Kirsch, Christian Rheinbay, and Roland Schnaubelt for clarifying discussions. Open access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 258734477 SFB 1173.

ORCID

Andreas Rieder https://orcid.org/0000-0002-3192-2847

REFERENCES

1. Haber E, Ascher UM. Preconditioned all-at-once methods for large, sparse parameter estimation problems. *Inverse Probl*. 2001;17:1847-1864. https://doi.org/10.1088/0266-5611/17/6/319
2. Haber E, Ascher UM, Oldenburg DW. Inversion of 3d electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*. 2004;69:1216-1228. https://doi.org/10.1190/1.1801938
3. Shenoy A, Heinkenschloss M, Cliff EM. Airfoil design by an all-at-once method. *Int J Comp Fluid Dyn*. 1998;11:3-25. https://doi.org/10.1080/10618569808940863
4. Kaltenbacher B. Regularization based on all-at-once formulations for inverse problems. *SIAM J Numer Anal*. 2016;54:2594-2618. https://doi.org/10.1137/16M1060984
5. Kaltenbacher B. All-at-once versus reduced iterative methods for time dependent inverse problems. *Inverse Problems*. 2017;33(4002):31. https://doi.org/10.1088/1361-6420/aa6f34
6. Fichtner A. Full Seismic Waveform Modelling and Inversion, Advances in Geophysical and Environmental Mechanics and Mathematics. 2011. https://doi.org/10.1007/978-3-642-15807-0
7. Virieux J, Operto S. An overview of full-waveform inversion in exploration geophysics. *Geophysics*. 2009;74:1-26. https://doi.org/10.1190/1.3238367
8. van Leeuwen T, Herrmann FJ. Mitigating local minima in full-waveform inversion by expanding the search space. *Geophys J Int*. 2013;195:661-667. https://doi.org/10.1093/gji/ggt258
9. van Leeuwen T, Herrmann FJ. A penalty method for PDE-constrained optimization in inverse problems. *Inverse Probl*. 2016;32(015007):26. https://doi.org/10.1088/0266-5611/32/1/015007
10. Aghamiry HS, Gholami A, Operto S. Compound regularization of full-waveform inversion for imaging piecewise media. *IEEE Trans Geosci Remote Sensing*. 2020;58:1192-1204. https://doi.org/10.1109/TGRS.2019.2944464

11. Kirsch A, Rieder A. Inverse problems for abstract evolution equations II: higher order differentiability for viscoelasticity. *SIAM J Appl Math*. 2019;79:2639-2662. https://doi.org/10.1137/19M1269403
12. Zeltmann U. The Viscoelastic Seismic Model: Existence, Uniqueness and Differentiability with Respect to Parameters. *PhD thesis*: Karlsruhe Institute of Technology; 2018. https://doi.org/10.5445/IR/1000093989
13. Blanch JO, Robertsson JOA, Symes WW. Modeling of a constant Q: methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. *Geophysics*. 1995;60:176-184. https://doi.org/10.1190/1.1443744
14. Bohlen T. Viskoelastische FD-Modellierung seismischer Wellen zur Interpretation gemessener Seismogramme. *PhD thesis*: Christian-Albrechts-Universität zu Kiel; 1998. https://bit.ly/2LM0SWr
15. Bohlen T. Parallel 3-D viscoelastic finite difference seismic modelling. *Comput Geosci*. 2002;28:887-899. https://doi.org/10.1016/S0098-3004(02)00006-7
16. Bohlen T, Fernandez MR, Ernesti J, Rheinbay C, Rieder A, Wieners C. 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
17. Kirsch A, Rieder A. Inverse problems for abstract evolution equations with applications in electrodynamics and elasticity. *Inverse Probl*. 2016;32(85001):24. https://doi.org/10.1088/0266-5611/32/8/085001
18. Schnaubelt R. Evolution equations, lecture notes, Karlsruhe Institute of Technology. 2019. https://www.math.kit.edu/iana3/schnaubelt/media/nevgl-skript.pdf
19. Engel K-J, Nagel R. *One-parameter semigroups for linear evolution equations, vol. 194 of Graduate Texts in Mathematics*. New York: Springer-Verlag; 2000. https://doi.org/10.1007/b97696
20. Hofmann B. 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. https://www.qucosa.de/fileadmin/data/qucosa/documents/4197/data/a008.pdf

How to cite this article: Rieder A. An all-at-once approach to full waveform inversion in the viscoelastic regime. *Math Meth Appl Sci*. 2021;1-13. https://doi.org/10.1002/mma.7190