Geophys. J. Int. (2016) **207**, 129-149

A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium

Pengliang Yang[1], Romain Brossier[1], Ludovic Métivier[2,1], Jean Virieux[1]
[1] *ISTerre, Univ. Grenoble Alpes, Grenoble, France. E-mail: pengliang.yang@univ-grenoble-alpes.fr*
[2] *Laboratoire Jean Kuntzmann (LJK), Univ. Grenoble Alpes, CNRS, Grenoble, France.*

Accepted 2016 July 13. Received 2016 July 13; in original form 2016 April 1

SUMMARY

In this paper we study 3D multiparameter full waveform inversion (FWI) in viscoelastic media based on the generalized Maxwell/Zener body (GMB/GZB) including arbitrary number of attenuation mechanisms. We present a frequency-domain energy analysis to establish the stability condition of a full anisotropic viscoelastic system, according to zero-valued boundary condition and the elastic-viscoelastic correspondence principle: the real-valued stiffness matrix becomes a complex-valued one in Fourier domain when seismic attenuation is taken into account. We develop a least-squares optimization approach to linearly relate the quality factor with the anelastic coefficients by estimating a set of constants which are independent of the spatial coordinates, which supplies an explicit incorporation of the parameter Q in the general viscoelastic wave equation. By introducing the Lagrangian multipliers into the matrix expression of the wave equation with implicit time integration, we build a systematic formulation of multiparameter full waveform inversion for full anisotropic viscoelastic wave equation, while the equivalent form of the state and adjoint equation with explicit time integration is available to be resolved efficiently. In particular, this formulation lays the foundation for the inversion of the parameter Q in the time domain with full anisotropic viscoelastic properties. In the 3D isotropic viscoelastic settings, the anelastic coefficients and the quality factors using bulk and shear moduli parametrization can be related to the counterparts using P- and S- velocity. Gradients with respect to any other parameter of interest can be found by chain rule. Pioneering numerical validations as well as the real applications of this most generic framework will be carried out to disclose the potential of viscoelastic FWI when adequate high performance computing resources and the field data are available.

Key words: Fourier analysis; Inverse theory; Seismic attenuation; Seismic tomography; Wave propagation

1 INTRODUCTION

Full waveform inversion (FWI) is an attractive tool to obtain high resolution subsurface parameters in complex geological structures, by iteratively minimizing the waveform misfit between the synthetic data and the observed seismograms (Tarantola 1987; Virieux & Operto 2009). The model parameters are therefore refined with the gradient of the objective function based on the adjoint method (Lailly 1983; Tarantola 1984; Plessix 2006), by simultaneously accessing to the forward wavefield and the adjoint wavefield. Due to the expensive computational cost, the significant advance of computer power in recent years has been made to promote FWI as a feasible and essential technology to build a good subsurface model in seismology and applied geophysics as long as seismic data provides the information. Although originally proposed by Tarantola (1984) in the time domain, the multiscale variants of FWI was developed in the other domains, such as frequency domain FWI (Pratt et al. 1998; Sirgue & Pratt 2004; Sirgue 2006; Operto et al. 2015), the Laplace domain (Shin & Cha 2009) and the Laplace-Fourier counterpart (Shin & Cha 2008). FWI has been explored in acoustic media (Gauthier et al. 1986) and elastic media (Mora 1987; Brossier et al. 2009; Vigh et al. 2014) from 2D to 3D (Vigh & Starr 2008), from single parameter (Operto et al. 2015) to multiple parameters (Prieux et al. 2013b,a; Zhou et al. 2015) in the time or frequency domain, allowing for isotropic and anisotropic propagation (Operto et al. 2009; Prieux et al. 2011; Gholami et al. 2013), based on different hardware architectures (Shin et al. 2014; Yang et al. 2015; Gokhberg & Fichtner 2015, just to name a few).

Seismic attenuation has already been well confirmed by a wide range of experimental tests and field observations. It plays a crucial role to delineate the absorption of the wave energy and the dispersion distortion for the phase of the waves, which have a strong impact on the FWI inversion result. There exists a number of rheological models (Ursin & Toverud 2002) to achieve attenuation, such as the power-law attenuation model (Strick 1967; Azimi et al. 1968; Kjartansson 1979), the Kolsky-Futterman model (Kolsky 1956; Futterman 1962), the

Maxwell body, the Kelvin-Voigt model (Casula & Carcione 1992), the Zener body or standard linear solid (SLS) (Ben-Menahem & Singh 1981; Pipkin 1986) (well known in mechanics and polymer chemistry (Alfrey 1948; Ferry 1961) and introduced to geophysics later), as well as their generalizations, the generalized Maxwell body (GMB-EK, abbreviated as GMB hereafter) (Emmerich & Korn 1987; Moczo & Kristek 2005) or its equivalence–the generalized Zener body (GZB) (Carcione et al. 1988b,c). The internal friction in the Earth has been well recognized to be nearly constant over a wide range of frequency band (Caputo 1967; Liu et al. 1976; Kjartansson 1979). In order to simulate attenuation for time-domain modeling, GMB/GZB have been widely used by researchers by the superposition of several mechanisms, which are physically meaningful to achieve the near constant Q effect to mimic the dissipation of the waves in the real Earth.

Recently, incorporating seismic attenuation in the framework of full waveform inversion (FWI) becomes an important topic in seismics. In the frequency domain, it is very convenient to incorporate attenuation by adding the parameter Q into the imaginary part of complex-valued velocity (Hicks & Pratt 2001). Inversion strategies for viscoacoustic waveform inversion have been proposed by Kamei & Pratt (2013). Operto et al. (2015) show that a 3D mono-parameter inversion in the frequency domain is feasible thanks to the progress of multifrontal solver for matrix inversion and the advance of computer capability. However, in the time domain only few studies investigate FWI in the presence of attenuation. Using a single attenuation mechanism, Bai et al. (2014) studied the viscoacoustic waveform inversion problem, while Cheng et al. (2015) did a viscoacoustic inversion for the parameter Q. Prior to their works, Kurzmann et al. (2013) have studied the impact of attenuation in time-domain viscoacoustic FWI using several attenuation mechanisms, showing that considering attenuation as a smooth background modeling parameter significantly improves the velocity reconstruction, while attenuation is considered in modelling and not an inversion parameter.

A general formulation has been provided by Tromp et al. (2005) for inverting subsurface parameters using adjoint formulation in viscoelastic media. The attenuation for shear velocity was emphasized due to its significance in the global scale wave propagation. Since the pioneer work of Romanowicz (1995) at global scale, estimation of the attenuation factor has focused the attention of seismologists (see Romanowicz & Mitchell 2007, for a review). Both the theory and the practice on multiparameter inversion using an arbitrary number of SLS mechanisms allows the inversion of the parameter Q in the time domain in full anisotropic viscoelastic medium: Fichtner & van Driel (2014) have provided simple expressions for frequency (in)-dependent Q models with numerical demonstration from regional- and global-scale time-domain wave propagation, complementing and simplifying nicely expressions provided by Tromp et al. (2005). In spite of this now well-defined framework, it is not so easy to find in the literature consistent expressions to be used for multiple parameters inversion of seismic waveforms including anisotropic parameters, density and attenuation factors. This is the purpose of this paper where expressions are presented in an explicit and coherent framework, allowing to clarify some implicit assumptions in previous demonstrations.

In this paper, we present a systematic formulation of multiparameter anisotropic viscoelastic FWI. Within the framework of linear viscoelasticity in Section 2, we perform an energy analysis in Section 3, based upon the elastic-viscoelastic correspondence principle, to disclose the requirement of GMB-based viscoelastic system to attenuate the energy in wave propagation. In Section 4, by reformulating the least-squares optimization to an approximate constant Q, we construct a linear representation of anelastic coefficients using attenuation parameters, which allows us to explicitly inject it into the viscoelastic wave equation and to invert this parameter further. In Section 5, we present the Lagrangian formulation of the waveform inversion problem with the matrix expression of symmetrical elastodynamic equation in an implicit time scheme, which provides the same solution as the equivalent non-symmetrical wave equation in the usual explicit time scheme for computation feasability. In Section 6, we establish the relationship between different model parametrization to facilitate the gradient computation of the misfit function between observed and synthetic fields, gradient with respect to bulk and shear moduli, as well as P- and S- wavespeeds. This new formulation builds a consistent framework with potential applications in both seismology and exploration geophysics.

2 LINEAR VISCOELASTICITY

For notation clarification, let us remind the usual framework of linear viscoelasticity by considering the partial differential equations before the correspondence principle between elastic and visco-elastic rheologies as well as the widely used generalized Maxwell/Zener body rheology.

The equation of wave motion is governed by the Newton's law

$$\rho \dot{v}_i = \sigma_{ij,j} = \partial_j \sigma_{ij}, \tag{1}$$

where the particle velocity v_i is related to the stress σ_{ij} scaled by the density ρ. In non-attenuating medium, the constitutive Hooke's law establishes the relationship between the stress σ_{ij} and the strain ϵ_{kl} by introducing the medium properties c_{ijkl} through the linear relation

$$\sigma_{ij} = c_{ijkl}\epsilon_{kl}, \tag{2}$$

where the Einstein notation has been tacitly applied (implicit summation for repeated indices $i, j, k, l \in [x, y, z] \equiv [1, 2, 3]$) in the remainder of this paper. The time derivatives are denoted by a dot over the involved variable for compactness. The strain ϵ is connected to the displacement u via

$$\epsilon_{ij} = \frac{1}{2}(u_{j,i} + u_{i,j}). \tag{3}$$

Thus, the time derivative of deformation is linked to the particle velocity $\mathbf{v} = (v_1, v_2, v_3)^T$ via $\dot{\epsilon}_{ij} = \frac{1}{2}(v_{j,i} + v_{i,j})$. Let us remark that usually the density and the stiffness are functions of space ($\rho = \rho(\mathbf{x})$, $c_{ijkl} = c_{ijkl}(\mathbf{x})$) while the particle velocity, the stress and the strain are functions of both space and time ($v_i = v_i(\mathbf{x}, t)$, $\sigma_{ij} = \sigma_{ij}(\mathbf{x}, t)$, $\epsilon_{kl} = \epsilon_{kl}(\mathbf{x}, t)$) when considering seismic wave propagation.

In the linear viscoelastic medium, the stress tensor σ_{ij} is generally related to the strain tensor ϵ_{ij} and the fourth-order tensorial relaxation function ψ_{ijkl} via

$$\sigma_{ij}(\mathbf{x},t) = \psi_{ijkl}(\mathbf{x},t) *_t \dot{\epsilon}_{kl}(\mathbf{x},t) = \int_{-\infty}^{t} \psi_{ijkl}(\mathbf{x},t-\tau)\dot{\epsilon}_{kl}(\mathbf{x},\tau)\mathrm{d}\tau \tag{4}$$

$$= \dot{\psi}_{ijkl}(\mathbf{x},t) *_t \epsilon_{kl}(\mathbf{x},t) = \int_{-\infty}^{t} \dot{\psi}_{ijkl}(\mathbf{x},t-\tau)\epsilon_{kl}(\mathbf{x},\tau)\mathrm{d}\tau. \tag{5}$$

The symbol $*_t$ stands for convolution in time satisfying the property $\partial_t(f *_t g) = (\partial_t f) *_t g = f *_t (\partial_t g)$. Let us denote the relaxation rate $M_{ijkl}(\mathbf{x},t) := \dot{\psi}_{ijkl}(\mathbf{x},t)$. This constitutive relation can be written as

$$\sigma_{ij}(\mathbf{x},t) = M_{ijkl}(\mathbf{x},t) *_t \epsilon_{kl}(\mathbf{x},t) \tag{6}$$

This yields, in the frequency domain (Emmerich & Korn 1987, eq. 1),

$$\tilde{\sigma}_{ij}(\mathbf{x},\omega) = \tilde{M}_{ijkl}(\mathbf{x},\omega)\tilde{\epsilon}_{kl}(\mathbf{x},\omega). \tag{7}$$

where the tilde over variables indicates the Fourier transform. Recall that the constitutive relation for pure elastic medium shown in (2), the pure elastic wave equation can be seen as a particular case of viscoelastic wave equation with specific relaxation rate

$$M_{ijkl}(\mathbf{x},t) = c_{ijkl}(\mathbf{x})\delta(t) \quad \text{or} \quad \tilde{M}_{ijkl}(\mathbf{x},\omega) = c_{ijkl}(\mathbf{x}). \tag{8}$$

where $\delta(t)$ is the Dirac delta function. Hence, $\tilde{M}_{ijkl}(\mathbf{x},\omega)$ is complex and frequency dependent for viscoelastic media, while it becomes real and frequency independent for elastic media. Equation (7) is the so-called elastic-viscoelastic correspondence principle, also referred to as the elastic-viscoelastic analogy, showing the similarity between the elastic and viscoelastic formulations in the frequency domain: the stiffness tensor becomes complex in viscous elastic media while keeping the same form (convolution in time becomes product in frequency). For more extensive explanations, see the contributions by Sips (1951); Read Jr (1950); Brull (1953); Lee (1960), Christensen (1982, p. 46) and Carcione (2001, p. 55). This very important fundamental principle is described in the frequency domain and, therefore, we shall consider the energy analysis in this domain in the following paragraph.

3 ENERGY ANALYSIS

Understanding what is the energy evolution in an attenuating medium is crucial for tracking reverse propagation of incident field for FWI aside better physical insight (Yang et al. 2016). In this section we show how to retrieve the decreasing rate of dissipation of the energy of the visco-elastodynamic equation based on GMB model. Previous equivalent analysis for GZB model has been performed in the time domain with a physically-based energy definition (Bécache et al. 2005). Through the frequency approach, we shall identify the term related to the energy definition as the explicit time-varying term, thanks to the correspondence principle. Finally, let us remark that the energy definition for general visco-elastodynamic models (not for very specific GMB or equivalent GZB models) is not unique as discussed by Scott (1997); Carcione (2001); Červenỳ & Pšenčík (2006).

We first define the inner product between two functions h and g over the spatial domain Ω as

$$\langle h, g\rangle_\Omega = \int_\Omega \mathrm{d}\mathbf{x} h(\mathbf{x},t)g(\mathbf{x},t). \tag{9}$$

Assuming zero-valued boundary condition for functions h and g (in our case, it will be velocities and stresses for wave propagation), the following fundamental relation can be deduced by integration by parts

$$\langle h, \partial_i g\rangle + \langle \partial_i h, g\rangle_\Omega = \int_\Omega \mathrm{d}\mathbf{x} h\partial_i g + \int_\Omega \mathrm{d}\mathbf{x}\partial_i h g = 0, \quad h \in \{v_i\}, g \in \{\sigma_{ij}\}, \quad i,j \in \{1,2,3\}. \tag{10}$$

The Fourier transform states that time domain multiplication corresponds to frequency domain convolution, leading to

$$\int_\Omega \mathrm{d}\mathbf{x}\tilde{h}^\dagger *_\omega \partial_i \tilde{g} + \int_\Omega \mathrm{d}\mathbf{x}(\partial_i \tilde{h})^\dagger *_\omega \tilde{g} = 0, \quad \tilde{h} \in \{\tilde{v}_i\}, \tilde{g} \in \{\tilde{\sigma}_{ij}\}, \quad i,j \in \{1,2,3\}, \tag{11}$$

where the symbol $*_\omega$ indicates frequency domain convolution, while the symbol † denotes conjugate transposition (equivalent to transpose ^{T} for real variables, and complex conjugate for scalars). When applied to operators, it will provide associated adjoint operator.

Let us define

$$D_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, D_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} D_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \tag{12}$$

such that a differential operator D with Auld's notation (Auld 1990) can be specified in the following

$$D = \begin{bmatrix} \partial_1 & 0 & 0 & 0 & \partial_3 & \partial_2 \\ 0 & \partial_2 & 0 & \partial_3 & 0 & \partial_1 \\ 0 & 0 & \partial_3 & \partial_2 & \partial_1 & 0 \end{bmatrix} = D_i\partial_i, \tag{13}$$

The time derivative of the strain tensor is

$$\dot{\epsilon}_{ij} = \frac{1}{2}(\dot{u}_{i,j} + \dot{u}_{j,i}) = \frac{1}{2}(v_{i,j} + v_{j,i}), \tag{14}$$

or in frequency domain

$$i\omega\tilde{\epsilon}_{ij}(\mathbf{x},\omega) = \frac{1}{2}(\partial_j \tilde{v}_i(\mathbf{x},\omega) + \partial_i \tilde{v}_j(\mathbf{x},\omega)). \tag{15}$$

The Newton law (1) in frequency domain reads

$$i\omega\rho(\mathbf{x})\tilde{v}_i(\mathbf{x},\omega) = \partial_j \tilde{\sigma}_{ij}(\mathbf{x},\omega), \tag{16}$$

which may be recast in matrix as

$$i\omega\rho \underbrace{\begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \tilde{v}_3 \end{bmatrix}}_{\tilde{\mathbf{v}}} = \underbrace{\begin{bmatrix} \partial_1 & 0 & 0 & 0 & \partial_3 & \partial_2 \\ 0 & \partial_2 & 0 & \partial_3 & 0 & \partial_1 \\ 0 & 0 & \partial_3 & \partial_2 & \partial_1 & 0 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \tilde{\sigma}_{11} \\ \tilde{\sigma}_{22} \\ \tilde{\sigma}_{33} \\ \tilde{\sigma}_{23} \\ \tilde{\sigma}_{13} \\ \tilde{\sigma}_{12} \end{bmatrix}}_{\tilde{\boldsymbol{\sigma}}}. \tag{17}$$

For the sake of simplicity we shall only keep the tilde over the variables and drop the index $(\mathbf{x},\omega)$ hereafter.

Instead of using complex-valued stiffness $\tilde{M}_{ijkl}$, we may also express the constitutive relation (7) by its inverse, the complex-valued compliance tensor $\tilde{S}$ through $\tilde{s}_{klij}\tilde{\sigma}_{ij} = \tilde{\epsilon}_{kl}$, leading to

$$i\omega\tilde{s}_{klij}\tilde{\sigma}_{ij} = i\omega\tilde{\epsilon}_{kl} = \frac{1}{2}(\partial_k \tilde{v}_l + \partial_l \tilde{v}_k). \tag{18}$$

Thanks to the symmetry assumption $\tilde{s}_{ijkl} = \tilde{s}_{jikl} = \tilde{s}_{ijlk} = \tilde{s}_{klij}$, we then have

$$i\omega \underbrace{\begin{bmatrix} \tilde{s}_{1111} & \tilde{s}_{1122} & \tilde{s}_{1133} & 2\tilde{s}_{1123} & 2\tilde{s}_{1113} & 2\tilde{s}_{1112} \\ \tilde{s}_{2211} & \tilde{s}_{2222} & \tilde{s}_{2233} & 2\tilde{s}_{2223} & 2\tilde{s}_{2213} & 2\tilde{s}_{2212} \\ \tilde{s}_{3311} & \tilde{s}_{3322} & \tilde{s}_{3333} & 2\tilde{s}_{3323} & 2\tilde{s}_{3313} & 2\tilde{s}_{3312} \\ 2\tilde{s}_{2311} & 2\tilde{s}_{2322} & 2\tilde{s}_{2333} & 4\tilde{s}_{2323} & 4\tilde{s}_{2313} & 4\tilde{s}_{2312} \\ 2\tilde{s}_{1311} & 2\tilde{s}_{1322} & 2\tilde{s}_{1333} & 4\tilde{s}_{1323} & 4\tilde{s}_{1313} & 4\tilde{s}_{1312} \\ 2\tilde{s}_{1211} & 2\tilde{s}_{1222} & 2\tilde{s}_{1233} & 4\tilde{s}_{1223} & 4\tilde{s}_{1213} & 4\tilde{s}_{1212} \end{bmatrix}}_{\tilde{S}} \underbrace{\begin{bmatrix} \tilde{\sigma}_{11} \\ \tilde{\sigma}_{22} \\ \tilde{\sigma}_{33} \\ \tilde{\sigma}_{23} \\ \tilde{\sigma}_{13} \\ \tilde{\sigma}_{12} \end{bmatrix}}_{\tilde{\boldsymbol{\sigma}}} = \underbrace{\begin{bmatrix} \partial_1 & 0 & 0 \\ 0 & \partial_2 & 0 \\ 0 & 0 & \partial_3 \\ 0 & \partial_1 & \partial_2 \\ \partial_3 & 0 & \partial_1 \\ \partial_2 & \partial_1 & 0 \end{bmatrix}}_{D^T} \underbrace{\begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \tilde{v}_3 \end{bmatrix}}_{\mathbf{v}}. \tag{19}$$

Combining equations (17) and (19) gives

$$i\omega \underbrace{\begin{bmatrix} \rho I & 0 \\ 0 & \tilde{S} \end{bmatrix}}_{\underline{S}} \underbrace{\begin{bmatrix} \tilde{\mathbf{v}} \\ \tilde{\boldsymbol{\sigma}} \end{bmatrix}}_{\tilde{\boldsymbol{w}}} = \underbrace{\begin{bmatrix} 0 & D \\ D^T & 0 \end{bmatrix}}_{B(\nabla)} \underbrace{\begin{bmatrix} \tilde{\mathbf{v}} \\ \tilde{\boldsymbol{\sigma}} \end{bmatrix}}_{\tilde{\boldsymbol{w}}}, B(\nabla) = \partial_i B_i, B_i = \begin{bmatrix} 0 & D_i \\ D_i^T & 0 \end{bmatrix}, \tag{20}$$

or, in compact form,

$$i\omega\underline{S}\tilde{\boldsymbol{w}} = B(\nabla)\tilde{\boldsymbol{w}}. \tag{21}$$

Convolving with the vector $\frac{1}{2}\tilde{\boldsymbol{w}}^\dagger$ on both sides of equation (21) gives

$$\begin{aligned} \frac{1}{2}\tilde{\boldsymbol{w}}^\dagger *_\omega i\omega\underline{S}\tilde{\boldsymbol{w}} =& \frac{1}{2}\tilde{\boldsymbol{w}}^\dagger *_\omega B(\nabla)\tilde{\boldsymbol{w}} = \frac{1}{2}\partial_i(\tilde{\boldsymbol{w}}^\dagger *_\omega B_i\tilde{\boldsymbol{w}}) \\ =& \frac{1}{2}\partial_i(\left[\tilde{\mathbf{v}}^\dagger, \tilde{\sigma}^\dagger\right] *_\omega \begin{bmatrix} 0 & D_i \\ D_i^T & 0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{v}} \\ \tilde{\sigma} \end{bmatrix}) = \frac{1}{2}\partial_i[\tilde{\mathbf{v}}^\dagger *_\omega D_i\tilde{\sigma} + \tilde{\sigma}^\dagger * D_i^T\tilde{\mathbf{v}}] \\ =& \Re[\partial_i(\tilde{\sigma}_{ij}^\dagger *_\omega \tilde{v}_j)] = \Re[\tilde{\sigma}_{ij,i}^\dagger *_\omega \tilde{v}_j + \tilde{\sigma}_{ij}^\dagger *_\omega \tilde{v}_{j,i}] \end{aligned} \tag{22}$$

where we have used the following identities

$$\rho\tilde{\mathbf{v}}^\dagger *_\omega \tilde{\mathbf{v}} = \rho\tilde{v}_i^\dagger *_\omega \tilde{v}_i, \quad \tilde{\sigma}^\dagger *_\omega i\omega S\tilde{\sigma} = (\tilde{M}_{ijkl}\tilde{\epsilon}_{ij})^\dagger *_\omega i\omega\tilde{\epsilon}_{kl} = (\tilde{M}\tilde{\epsilon})^\dagger *_\omega i\omega\tilde{\epsilon}. \tag{23}$$

Also note that

$$\frac{1}{2}\tilde{\boldsymbol{w}}^\dagger *_\omega i\omega\underline{S}\tilde{\boldsymbol{w}} = \frac{1}{2}(\tilde{\mathbf{v}}^\dagger *_\omega i\omega\rho\tilde{\mathbf{v}} + \tilde{\sigma}^\dagger *_\omega i\omega S\tilde{\sigma}). \tag{24}$$

Combining equation (24) with (22) yields

$$\frac{1}{2}(\tilde{\mathbf{v}}^\dagger *_\omega i\omega\rho\tilde{\mathbf{v}} + \tilde{\sigma}^\dagger *_\omega i\omega S\tilde{\sigma}) = \Re[\tilde{\sigma}_{ij,i}^\dagger *_\omega \tilde{v}_j + \tilde{\sigma}_{ij}^\dagger *_\omega \tilde{v}_{j,i}]. \tag{25}$$

By integration over the whole domain Ω, we obtain

$$\frac{1}{2}\int_\Omega \mathrm{d}\mathbf{x}(\rho\tilde{\mathbf{v}}^\dagger *_\omega i\omega\tilde{\mathbf{v}} + (\tilde{M}\tilde{\epsilon})^\dagger *_\omega i\omega\tilde{\epsilon}) = \Re\int_\Omega \mathrm{d}\mathbf{x}[\tilde{\sigma}_{ij,i}^\dagger *_\omega \tilde{v}_j + \tilde{\sigma}_{ij}^\dagger *_\omega \tilde{v}_{j,i}] = 0, \tag{26}$$

thanks to the identity (11).

Let us consider GMB/GZB using L SLSs mechanisms associated with complex stiffness tensor $\tilde{M}_{ijkl} = c_{ijkl}(1-\sum_{\ell=1}^L Y_\ell^{ijkl}\omega_\ell/(\omega_\ell+i\omega))$ (we shall give further detail on it in the next sections and in the Appendix A). We have

$$\begin{aligned}
&(\tilde{M}_{ijkl}\tilde{\epsilon}_{ij})^\dagger *_\omega i\omega\tilde{\epsilon}_{kl} = c_{ijkl}\left((1-\sum_{\ell=1}^L Y_\ell^{ijkl}\frac{\omega_\ell}{\omega_\ell+i\omega})\tilde{\epsilon}_{ij}\right)^\dagger *_\omega i\omega\tilde{\epsilon}_{kl}\\
=&c_{ijkl}(1-\sum_{\ell=1}^L Y_\ell^{ijkl})\tilde{\epsilon}_{ij}^\dagger *_\omega i\omega\tilde{\epsilon}_{kl} + \sum_{\ell=1}^L Y_\ell^{ijkl}c_{ijkl}\left(\frac{1}{\omega_\ell}\tilde{\xi}_\ell^{ij}\right)^\dagger *_\omega \left(\frac{1}{\omega_\ell}\tilde{\xi}_\ell^{kl}\right)(i\omega+\omega_\ell) \quad (\frac{1}{\omega_\ell}\tilde{\xi}_\ell^{ij} := \frac{i\omega}{i\omega+\omega_\ell}\tilde{\epsilon}_{ij})\\
=&c_{ijkl}(1-\sum_{\ell=1}^L Y_\ell^{ijkl})\tilde{\epsilon}_{ij}^\dagger *_\omega i\omega\tilde{\epsilon}_{kl} + \sum_{\ell=1}^L \frac{Y_\ell^{ijkl}}{\omega_\ell^2}c_{ijkl}\tilde{\xi}_\ell^{ij\dagger} *_\omega i\omega\tilde{\xi}_\ell^{kl} + \sum_{\ell=1}^L \frac{Y_\ell^{ijkl}}{\omega_\ell}c_{ijkl}\tilde{\xi}_\ell^{ij\dagger} *_\omega \tilde{\xi}_\ell^{kl}
\end{aligned} \tag{27}$$

where we remind that the memory variables ξ_ℓ^{ij} satisfies the equation $\partial_t\xi_\ell^{ij} + \omega_\ell\xi_\ell^{ij} = \omega_\ell\epsilon_{ij}$. In terms of (26), we obtain the important visco-elastic/elastic identity:

$$\begin{aligned}
0 =&\frac{1}{2}\int_\Omega \mathrm{d}\mathbf{x}(\rho\tilde{\mathbf{v}}^\dagger *_\omega i\omega\tilde{\mathbf{v}} + (\tilde{M}\tilde{\epsilon})^\dagger *_\omega i\omega\tilde{\epsilon})\\
=&\frac{1}{2}\int_\Omega \mathrm{d}\mathbf{x}\left(\rho\tilde{v}_i^\dagger *_\omega i\omega\tilde{v}_i + c_{ijkl}(1-\sum_{\ell=1}^L Y_\ell^{ijkl})\tilde{\epsilon}_{ij}^\dagger *_\omega i\omega\tilde{\epsilon}_{kl} + \sum_{\ell=1}^L \frac{Y_\ell^{ijkl}}{\omega_\ell^2}c_{ijkl}\tilde{\xi}_\ell^{ij\dagger} *_\omega i\omega\tilde{\xi}_\ell^{kl}\right)\\
&+\frac{1}{2}\sum_{\ell=1}^L\frac{1}{\omega_\ell}\int_\Omega \mathrm{d}\mathbf{x}Y_\ell^{ijkl}c_{ijkl}\tilde{\xi}_\ell^{ij\dagger} *_\omega \tilde{\xi}_\ell^{kl}.
\end{aligned} \tag{28}$$

Translating the identity (28) into time domain collecting all terms with $i\omega$ as time derivatives yields

$$\begin{aligned}
0 =&\frac{1}{2}\int_\Omega \mathrm{d}\mathbf{x}\Big(\rho v_i^\dagger\partial_t v_i + c_{ijkl}(1-\sum_{\ell=1}^L Y_\ell^{ijkl})\epsilon_{ij}^\dagger\partial_t\epsilon_{kl} + \sum_{\ell=1}^L \frac{Y_\ell^{ijkl}}{\omega_\ell^2}c_{ijkl}\xi_\ell^{ij\dagger}\partial_t\xi_\ell^{kl}\Big)\\
&+\frac{1}{2}\sum_{\ell=1}^L\frac{1}{\omega_\ell}\int_\Omega \mathrm{d}\mathbf{x}Y_\ell^{ijkl}c_{ijkl}\xi_\ell^{ij\dagger}\xi_\ell^{kl}\\
=&\frac{1}{2}\partial_t\underbrace{\left(\frac{1}{2}\int_\Omega \mathrm{d}\mathbf{x}\Big(\rho v_i^\dagger v_i + c_{ijkl}(1-\sum_{\ell=1}^L Y_\ell^{ijkl})\epsilon_{ij}^\dagger\epsilon_{kl} + \sum_{\ell=1}^L \frac{Y_\ell^{ijkl}}{\omega_\ell^2}c_{ijkl}\xi_\ell^{ij\dagger}\xi_\ell^{kl}\Big)\right)}_{E}\\
&+\frac{1}{2}\sum_{\ell=1}^L\frac{1}{\omega_\ell}\int_\Omega \mathrm{d}\mathbf{x}Y_\ell^{ijkl}c_{ijkl}\xi_\ell^{ij\dagger}\xi_\ell^{kl}
\end{aligned} \tag{29}$$

Equation (29) is very important as it reveals a balance of the total energy for the above GZB/GMB-based viscoelastic system defined as

$$E = \frac{1}{2}\int_\Omega \mathrm{d}\mathbf{x}\Big(\rho v_i^\dagger v_i + c_{ijkl}(1-\sum_{\ell=1}^L Y_\ell^{ijkl})\epsilon_{ij}^\dagger\epsilon_{kl} + \sum_{\ell=1}^L \frac{Y_\ell^{ijkl}}{\omega_\ell^2}c_{ijkl}\xi_\ell^{ij\dagger}\xi_\ell^{kl}\Big). \tag{30}$$

A sufficient condition for having a negative time derivative of time of the energy will require that all the 4-th order tensors $C :: Y_\ell$ could be associated to positive definite bilinear functions as shown in the following expression

$$\partial_t E = -\sum_{\ell=1}^L\frac{1}{\omega_\ell}\int_\Omega \mathrm{d}\mathbf{x}Y_\ell^{ijkl}c_{ijkl}\xi_\ell^{ij\dagger}\xi_\ell^{kl} < 0, \tag{31}$$

essential for dissipative viscoelastic media. Let us remark that Bécache et al. (2005) has expressed it globally over the sum as a necessary and sufficient condition.

As a conclusion, we have been able to show that the total energy, which is defined unambiguously for GZB/GMB models, is dissipated inside the media during propagation: it will be a key criterion for reverse propagation of the incident field when computing the gradient of the FWI (Yang et al. 2016). Note that the derivation of this energy balance can be done directly in the time domain for elastic and viscoelastic isotropic medium: see Appendix B for more details.

4 CONSTANT Q APPROXIMATION

The number of independent components in M_{ijkl} can be reduced to 21 due to the symmetry of the stress and strain tensors, and also to the unique strain energy definition for elastic materials (Carcione 2001, p. 55). For GMB/GZB rheology, this number of parameters still holds, thanks to the linearity. This reduction allows the introduction of the Voigt indexing, $(11) \to 1$, $(22) \to 2$, $(33) \to 3$, $(23) = (32) \to 4$, $(13) = (31) \to 5$, $(12) = (21) \to 6$, which permits us to consider stresses and strains as vectors and 4x4 stiffness tensors as 2x2 tensors (equivalent to a matrices) through relations in the time domain

$$\sigma_I(\mathbf{x},t) = M_{IJ}(\mathbf{x},t) *_t \epsilon_J(\mathbf{x},t), \quad I,J = 1,\cdots,6, \tag{32}$$

which is expressed in the frequency domain

$$\tilde{\sigma}_I(\mathbf{x},\omega) = \tilde{M}_{IJ}(\mathbf{x},\omega)\tilde{\epsilon}_J(\mathbf{x},\omega), \quad I,J=1,\cdots,6. \tag{33}$$

Let us remind that for $J=4,5,6$, the symmetric quantities collapse in $\tilde{\epsilon}_J(\mathbf{x},\omega)$. That is, $\tilde{\epsilon}_4=2\tilde{\epsilon}_{23}$, $\tilde{\epsilon}_5=2\tilde{\epsilon}_{13}$, $\tilde{\epsilon}_6=2\tilde{\epsilon}_{12}$ (Carcione & Cavallini 1993).

The seismic attenuation is usually expressed as the inverse of the quality factor Q (also called internal friction or dissipation factor) defined as the the ratio of real and imaginary parts of the complex-valued modulus $\tilde{M}_{IJ}(\mathbf{x},\omega)$

$$\tilde{Q}_{IJ}^{-1}(\mathbf{x},\omega) = \frac{\Im[\tilde{M}_{IJ}(\mathbf{x},\omega)]}{\Re[\tilde{M}_{IJ}(\mathbf{x},\omega)]}. \tag{34}$$

Numerous observations assess that nearly nearly constant Q assumption over a wide range of frequencies is valid in seismic wave propagation in the Earth (Caputo 1967; Kjartansson 1979). Fitting a constant Q parameter over this frequency range $\omega \in [\omega_{\min},\omega_{\max}]$ requires considering a number of attenuation mechanisms. Based upon GMB/GZB using L attenuation mechanisms (Emmerich & Korn 1987; Carcione et al. 1988a; Moczo & Kristek 2005; Moczo et al. 2007b), one may introduce a number of dimensionless anelastic coefficients Y_ℓ^{IJ} in the definition of $M_{IJ}(\mathbf{x},\omega)$ with a number of specific reference frequencies $\omega_\ell \in [\omega_{\min},\omega_{\max}]$, $\ell=1,\cdots,L$:

$$\tilde{M}_{IJ}(\mathbf{x},\omega) = c_{IJ}(\mathbf{x})\left(1-\sum_{\ell=1}^{L} Y_\ell^{IJ}(\mathbf{x})\frac{\omega_\ell}{\omega_\ell+i\omega}\right). \tag{35}$$

Inserting (35) into (33) leads to

$$\tilde{\sigma}_I(\mathbf{x},\omega) = c_{IJ}(\mathbf{x})\left(\tilde{\epsilon}_J(\mathbf{x},\omega)-\sum_{\ell=1}^{L} Y_\ell^{IJ}(\mathbf{x})\tilde{\zeta}_\ell^J(\mathbf{x},\omega)\right), \tilde{\zeta}_\ell^J(\mathbf{x},\omega) := \frac{\omega_\ell}{\omega_\ell+i\omega}\tilde{\epsilon}_J(\mathbf{x},\omega) \tag{36}$$

or in time domain

$$\sigma_I(\mathbf{x},t) = c_{IJ}(\mathbf{x})\left(\epsilon_J(\mathbf{x},t)-\sum_{\ell=1}^{L} Y_\ell^{IJ}(\mathbf{x})\zeta_\ell^J(\mathbf{x},t)\right), \partial_t\zeta_\ell^J(\mathbf{x},t)+\omega_\ell\zeta_\ell^J(\mathbf{x},t) = \omega_\ell\epsilon_J(\mathbf{x},t). \tag{37}$$

According to (34) and (35), the resulting quality factor for GMB/GZB is given by (Moczo et al. 2007a, eq. 117)

$$\tilde{Q}_{IJ}^{-1}(\mathbf{x},\omega) = \frac{\sum_{\ell=1}^{L} Y_\ell^{IJ}(\mathbf{x})\frac{\omega_\ell\omega}{\omega_\ell^2+\omega^2}}{1-\sum_{\ell=1}^{L} Y_\ell^{IJ}(\mathbf{x})\frac{\omega_\ell^2}{\omega_\ell^2+\omega^2}} \approx \sum_{\ell=1}^{L} Y_\ell^{IJ}(\mathbf{x})\frac{\omega_\ell\omega}{\omega^2+\omega_\ell^2}, \tag{38}$$

where we apply the approximation $1-\sum_{\ell=1}^{L} Y_\ell^{IJ}(\mathbf{x})\omega_\ell^2/(\omega^2+\omega_\ell^2)\approx 1$ under the assummption $\tilde{Q}_{IJ}(\mathbf{x},\omega)\gg 1$ in realistic attenuative media (Blanch et al. 1995). Equation (38) provides us an important linear relationship between $Q_{IJ}^{-1}(\mathbf{x},\omega)$ and the anelastic coefficients $Y_\ell^{IJ}(\mathbf{x})$.

For an approximation of a constant $Q_{IJ}(\mathbf{x})$ over the used frequency band $[\omega_{\min},\omega_{\max}]$, the $Y_\ell^{IJ}(\mathbf{x})$ are usually determined by minimizing the least-squares problems (Blanch et al. 1995; Bohlen 2002)

$$\min_{Y_\ell^{IJ}(\mathbf{x})} \chi_0^{Q_{IJ}(\mathbf{x})}, \quad \chi_0^{Q_{IJ}(\mathbf{x})} = \int_{\omega_{\min}}^{\omega_{\max}} (\tilde{Q}_{IJ}^{-1}(\mathbf{x},\omega)-Q_{IJ}^{-1}(\mathbf{x}))^2 \mathrm{d}\omega. \tag{39}$$

It is important to emphasize that $\tilde{Q}(\mathbf{x},\omega)$ is a frequency dependent quality factor coming from GMB/GZB model associated with equation (38), while $Q(\mathbf{x})$ is a frequency independent target value for attenuation (and also a parameter to be inverted through FWI iterations). Generally, these least-squares minimizations have to be performed for each spatial location $\mathbf{x}\in\Omega$ and each IJ component. Each minimization is providing the ℓ corresponding coefficients.

Thanks to the assumption of frequency independent $Q_{IJ}(\mathbf{x})$, the objective functions $\chi_0^{Q_{IJ}(\mathbf{x})}$ can be recast as

$$\chi_0^{Q_{IJ}(\mathbf{x})} = Q_{IJ}^{-2}(\mathbf{x})\gamma^2 \int_{\omega_{\min}}^{\omega_{\max}} (\sum_{\ell=1}^{L}\gamma_\ell^{IJ}(\mathbf{x})\frac{\omega_\ell\omega}{\omega^2+\omega_\ell^2}-\gamma^{-1})^2\mathrm{d}\omega \tag{40}$$

where a user-defined constant γ (usually chosen such that $\gamma\in[Q_{IJ}^{\min},Q_{IJ}^{\max}]$) and the new variables $\gamma_\ell^{IJ}(\mathbf{x}) := \gamma^{-1}Q_{IJ}(\mathbf{x})Y_\ell^{IJ}(\mathbf{x})$ have been introduced. Let us now define a new set of minimization problems to obtain $\gamma_\ell^{IJ}(\mathbf{x})$

$$\min_{\gamma_\ell^{IJ}(\mathbf{x})} \chi_1^{Q_{IJ}(\mathbf{x})}, \quad \chi_1^{Q_{IJ}(\mathbf{x})} = \int_{\omega_{\min}}^{\omega_{\max}} (\sum_{\ell=1}^{L}\gamma_\ell^{IJ}(\mathbf{x})\frac{\omega_\ell\omega}{\omega^2+\omega_\ell^2}-\gamma^{-1})^2\mathrm{d}\omega. \tag{41}$$

equivalent to equation (40) thanks to the relation $\chi_0^{Q_{IJ}(\mathbf{x})} = \gamma^2 Q_{IJ}^{-2}(\mathbf{x})\chi_1^{Q_{IJ}(\mathbf{x})}$ and $Y_\ell^{IJ}(\mathbf{x}) = \gamma\,\gamma_\ell^{IJ}(\mathbf{x})Q_{IJ}^{-1}(\mathbf{x})$.

A careful analysis of equation (41) shows that all these minimization problems are actually the same, resulting in the same space independent and component independent solutions $\gamma_\ell := \gamma_\ell^{IJ}(\mathbf{x})$. That means we only need to perform one least-squares optimization to obtain the anelastic coefficients specified by

$$Y_\ell^{IJ}(\mathbf{x}) = y_\ell Q_{IJ}^{-1}(\mathbf{x}) \text{ with } y_\ell = \gamma\,\gamma_\ell. \tag{42}$$

In other words, the previous reformulation of the least-squares minimizations, based on the frequency independant attenuation, helps us to establish an important separability relation between L constants y_ℓ and $Q_{IJ}(\mathbf{x})$ shown in (42). This has a significant impact for numerical implementation: instead of storing the $L \times 21$ anelastic coefficients $Y_\ell^{IJ}(\mathbf{x})$ at each spatial location $\mathbf{x}$, we only need to store L scalars y_ℓ and the space dependent model parameter $Q_{IJ}(\mathbf{x})$. The benefit of the use of single parameter γ to determine the magnitude of several mechanisms has also been emphasized by (Blanch et al. 1995), although the details were omitted. In addition, the attenuation/quality factor $Q_{IJ}(\mathbf{x})$ can be explicitly incorporated in the wave equation, and therefore naturally considered to be reconstructed in the FWI framework.

5 3D VISCOELASTIC INVERSION

5.1 General viscoelastic wave equation

Using the Voigt notation, we can write the time derivative of the stress as

$$\partial_t \sigma_I(\mathbf{x},t) = c_{IJ}(\mathbf{x}) \left(\dot{\epsilon}_J(\mathbf{x},t) - Q_{IJ}^{-1}(\mathbf{x}) \sum_{\ell=1}^{L} y_\ell \xi_\ell^J(\mathbf{x},t) \right). \tag{43}$$

For wave modeling, one prefers using stiffness tensor c_{IJ} than compliance tensor s_{IJ}. Let us remark that expressions (43) are very general as we do not enforce the isotropic assumption of attenuation as in (A.15) given by Kristek & Moczo (2003); Moczo et al. (2007a). Let us introduce an attenuation vector ξ_ℓ such that

$$\boldsymbol{\xi}_\ell = (\xi_\ell^1, \xi_\ell^2, \cdots, \xi_\ell^6)^T = (\xi_\ell^{11}, \xi_\ell^{22}, \xi_\ell^{33}, \xi_\ell^{23}, \xi_\ell^{13}, \xi_\ell^{12})^T. \tag{44}$$

We may write the memory variable ordinary differential equation in a system

$$\partial_t \boldsymbol{\xi}_\ell(\mathbf{x},t) + \omega_\ell \boldsymbol{\xi}_\ell(\mathbf{x},t) = \omega_\ell D^T \mathbf{v}(\mathbf{x},t). \tag{45}$$

Let us remind that again the last three subequations collect the relation for the memory variables $\xi_\ell^{23}, \xi_\ell^{13}, \xi_\ell^{12}$ with symmetric index. Putting all things together, the viscoelastic system including the external sources reads

$$\rho \partial_t \mathbf{v}(\mathbf{x},t) = D\boldsymbol{\sigma}(\mathbf{x},t) + \mathbf{f}_\mathbf{v}(\mathbf{x},t) \tag{46a}$$

$$\partial_t \boldsymbol{\sigma}(\mathbf{x},t) = CD^T \mathbf{v}(\mathbf{x},t) - (C :: \Gamma)(\mathbf{x}) \sum_{\ell=1}^{L} y_\ell \boldsymbol{\xi}_\ell(\mathbf{x},t) + \mathbf{f}_\sigma(\mathbf{x},t) \tag{46b}$$

$$\partial_t \boldsymbol{\xi}_\ell(\mathbf{x},t) + \omega_\ell \boldsymbol{\xi}_\ell(\mathbf{x},t) = \omega_\ell D^T \mathbf{v}(\mathbf{x},t), \quad \ell = 1, \cdots, L, \tag{46c}$$

where we remind the following definitions:

$$\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6)^T = (\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{23}, \sigma_{13}, \sigma_{12})^T \tag{47}$$

$$\mathbf{f}_\mathbf{v} = (f_{v_1}, f_{v_2}, f_{v_3}), \mathbf{f}_\sigma = (f_{\sigma_1}, f_{\sigma_2}, f_{\sigma_3}, f_{\sigma_4}, f_{\sigma_5}, f_{\sigma_6})^T \tag{48}$$

$$C = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} \end{bmatrix} \text{ and } \Gamma = \begin{bmatrix} Q_{11}^{-1} & Q_{12}^{-1} & Q_{13}^{-1} & Q_{14}^{-1} & Q_{15}^{-1} & Q_{16}^{-1} \\ Q_{21}^{-1} & Q_{22}^{-1} & Q_{23}^{-1} & Q_{24}^{-1} & Q_{25}^{-1} & Q_{26}^{-1} \\ Q_{31}^{-1} & Q_{32}^{-1} & Q_{33}^{-1} & Q_{34}^{-1} & Q_{35}^{-1} & Q_{36}^{-1} \\ Q_{41}^{-1} & Q_{42}^{-1} & Q_{43}^{-1} & Q_{44}^{-1} & Q_{45}^{-1} & Q_{46}^{-1} \\ Q_{51}^{-1} & Q_{52}^{-1} & Q_{53}^{-1} & Q_{54}^{-1} & Q_{55}^{-1} & Q_{56}^{-1} \\ Q_{61}^{-1} & Q_{62}^{-1} & Q_{63}^{-1} & Q_{64}^{-1} & Q_{65}^{-1} & Q_{66}^{-1} \end{bmatrix}.$$

Due to the existence of the compliance matrix $S = C^{-1}$, the equation (46) can be written under an implicit time partial differential equations as

$$\begin{aligned} &\underbrace{\begin{bmatrix} \rho I_3 & 0 & \cdots & 0 & \cdots \\ 0 & C^{-1} & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & 0 & \frac{1}{\omega_\ell} I_6 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{N_1(\mathbf{m})} \partial_t \underbrace{\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\sigma} \\ \vdots \\ \boldsymbol{\xi}_\ell \\ \vdots \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} 0 & D & \cdots & 0 & \cdots \\ D^T & 0 & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ D^T & \vdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{B_1(\nabla)} \underbrace{\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\sigma} \\ \vdots \\ \boldsymbol{\xi}_\ell \\ \vdots \end{bmatrix}}_{\mathbf{w}} \\ &\qquad - \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 & \cdots \\ 0 & 0 & \cdots & y_\ell C^{-1}(C :: \Gamma) & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & 0 & I_6 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{N_2(\mathbf{m})} \underbrace{\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\sigma} \\ \vdots \\ \boldsymbol{\xi}_\ell \\ \vdots \end{bmatrix}}_{\mathbf{w}} + \underbrace{\begin{bmatrix} \mathbf{f}_\mathbf{v} \\ C^{-1}\mathbf{f}_\sigma \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{s}}, \end{aligned} \tag{49}$$

Equation (49) can be summarized as

$$F(\mathbf{m},\mathbf{w}) = N_1(\mathbf{m})\partial_t\mathbf{w} - B_1(\nabla)\mathbf{w} + N_2(\mathbf{m})\mathbf{w} - \mathbf{s} = 0, \tag{50}$$

where the model parameters $\mathbf{m} = (\rho(\mathbf{x}), C_{IJ}(\mathbf{x}), Q_{IJ}^{-1}(\mathbf{x}))^\dagger$ are explicitly written in the matrices N_1, N_2. Please note that we have separated these matrices such that B_1 is merely related to spatial derivatives. Let us underline again that the system (50) is implicit in time as the matrix N_1 is not diagonal ($S = C^{-1}$ is not diagonal due to full anisotropy). As it is always possible to invert the matrix N_1, we go back to the equation (46) in a compact form as

$$\partial_t\mathbf{w} - N_1^{-1}(\mathbf{m})B_1(\nabla)\mathbf{w} + N_1^{-1}(\mathbf{m})N_2(\mathbf{m})\mathbf{w} - N_1^{-1}(\mathbf{m})\mathbf{s} = 0, \tag{51}$$

which can be solved using explicit time integration.

5.2 3D FWI in general viscoelastics

5.2.1 Preliminary

The inner product between h and g over spatial domain Ω and the time duration $[0, T]$ is defined as

$$\langle h, g\rangle_{\Omega\times T} = \int_0^T \mathrm{d}t \int_\Omega \mathrm{d}\mathbf{x} h(\mathbf{x},t)g(\mathbf{x},t) \tag{52}$$

Integrating by parts for temporal and spatial coordinates yields

$$\begin{aligned}\langle h, \partial_t g\rangle_{\Omega\times T} =& \int_0^T \mathrm{d}t \int_\Omega \mathrm{d}\mathbf{x} h(\mathbf{x},t)\partial_t g(\mathbf{x},t)\\ =& \left[\int_\Omega \mathrm{d}\mathbf{x} h(\mathbf{x},t)g(\mathbf{x},t)\right]_0^T - \int_0^T \mathrm{d}t \int_\Omega \mathrm{d}\mathbf{x} \partial_t h(\mathbf{x},t)g(\mathbf{x},t)\\ =& -\langle \partial_t h, g\rangle_{\Omega\times T}\end{aligned} \tag{53}$$

assuming the initial condition $g(\mathbf{x}, t) = 0$ and the finial condition $h(\mathbf{x}, T) = 0$, and

$$\begin{aligned}\langle h, \partial_i g\rangle_{\Omega\times T} =& \int_0^T \mathrm{d}t \int_\Omega \mathrm{d}\mathbf{x} h(\mathbf{x},t)\partial_i g(\mathbf{x},t)\\ =& \left[\int_0^T \mathrm{d}t h(\mathbf{x},t)g(\mathbf{x},t)\right]_{\partial\Omega} - \int_0^T \mathrm{d}t \int_\Omega \mathrm{d}\mathbf{x} \partial_i h(\mathbf{x},t)g(\mathbf{x},t)\\ =& -\langle \partial_i h, g\rangle_{\Omega\times T}, i \in \{x,y,z\}\end{aligned} \tag{54}$$

assuming the zero-valued boundary condition $h|_{\partial\Omega} = 0$ or $g|_{\partial\Omega} = 0$. Recall that the adjoint of the operator L, namely $L^\dagger$, is defined as $\langle h, Lg\rangle = \langle L^\dagger h, g\rangle$. Therefore, the equations (53) and (54) imply the following adjoint operators

$$(\partial_i)^\dagger = -\partial_i, (D)^\dagger = (D_i\partial_i)^\dagger = -D_i^T\partial_i = -D^T, (\partial_t)^\dagger = -\partial_t, \tag{55}$$

in the assumption of zero-valued boundary condition, as well as initial and final conditions.

5.2.2 Full waveform inversion

Full waveform inversion tries to reduce the data misfit between the synthetic and the observed data at the receiver coordinates by iteratively minimizing a least-squares objective functional:

$$\chi(\mathbf{m},\mathbf{w}) = \frac{1}{2}\|R_r\mathbf{w} - \mathbf{d}\|^2, \tag{56}$$

where $\mathbf{d} := d(\mathbf{x}_r, t)$ is the observed seismogram, and R_r a restriction operator mapping the full wavefield $\mathbf{w}(\mathbf{x}, t)$ onto receiver locations. Note that a summation over all excitation sources has been implied. Then we introduce the augmented Lagrangian functional with the Lagrangian multiplier $\bar{\mathbf{w}}$

$$\mathbb{L}(\mathbf{m},\mathbf{w},\bar{\mathbf{w}}) = \chi(\mathbf{m},\mathbf{w}) + \langle\bar{\mathbf{w}}, F(\mathbf{m},\mathbf{w})\rangle_{\Omega\times T}. \tag{57}$$

- Differentiation with respect to Lagrangian multiplier $\bar{\mathbf{w}}$ leads to the state wave equation (50).
- Differentiation with respect to state wavefield variable $\mathbf{w}$ gives the adjoint state equation expressed as

$$\frac{\partial\mathbb{L}}{\partial\mathbf{w}} = \frac{\partial\chi}{\partial\mathbf{w}} + \left(\frac{\partial F(\mathbf{m},\mathbf{w})}{\partial\mathbf{w}}\right)^\dagger \bar{\mathbf{w}} = 0 \tag{58}$$

$$\Leftrightarrow (N_1(\mathbf{m})\partial_t - B_1(\nabla) + N_2(\mathbf{m}))^\dagger\bar{\mathbf{w}} = -R_r^\dagger(R_r\mathbf{w} - \mathbf{d}), \tag{59}$$

or

$$N_1^\dagger(\mathbf{m})\partial_t\bar{\mathbf{w}} + B_1^\dagger(\nabla)\bar{\mathbf{w}} - N_2^\dagger(\mathbf{m})\bar{\mathbf{w}} = R_r^\dagger(R_r\mathbf{w} - \mathbf{d}), \tag{60}$$

where the Lagrangian multiplier $\bar{\mathbf{w}}$ is the so-called adjoint state vector, for which a final condition $\bar{\mathbf{w}}(\mathbf{x},T)=0$ has been employed. Let us remind again that the symbol † denotes the transpose conjugate operation which turns out to be the transpose one as we have real values in the time formulation we consider. The adjoint equation (60) involves a reverse time propagation with negative attenuation, which is numerically stable as a forward time propagation with positive attenuation (Tarantola 1988).

- The gradient of the misfit function with respect to model parameters $\mathbf{m}$ is the same as the Lagrangian at saddle points

$$\frac{\partial \chi}{\partial \mathbf{m}} = \frac{\partial \mathbb{L}}{\partial \mathbf{m}} = \langle \bar{\mathbf{w}}, \frac{\partial F(\mathbf{m},\mathbf{w})}{\partial \mathbf{m}} \rangle_T = \langle \bar{\mathbf{w}}, \frac{\partial F(\mathbf{m},\mathbf{w})}{\partial \mathbf{m}} \rangle_T \tag{61}$$

It is worth noting that the wave equation (50) and the adjoint equation (60) used in the Lagrangian formulation are implicit time integration systems. Multiplying $-N_1^{-\dagger}(\mathbf{m})$ on both sides for the equations (60) yields the expression of the adjoint using explicit time integration

$$\partial_t \bar{\mathbf{w}} + N_1^{-\dagger}(\mathbf{m}) B_1^\dagger(\nabla)\bar{\mathbf{w}} - N_1^{-\dagger}(\mathbf{m}) N_2^\dagger(\mathbf{m})\bar{\mathbf{w}} = N_1^{-\dagger}(\mathbf{m}) R_r^\dagger (R_r \mathbf{w} - \mathbf{d}) \tag{62}$$

where

$$N_1^{-\dagger}(\mathbf{m}) = N_1^{-1}(\mathbf{m}) = \begin{bmatrix} \frac{1}{\rho} I_3 & 0 & \cdots & 0 & \cdots \\ 0 & C & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & 0 & \omega_\ell I_6 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, B_1^\dagger(\nabla) = \begin{bmatrix} 0 & -D & \cdots & -D & \cdots \\ -D^T & 0 & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

and

$$N_1^{-\dagger}(\mathbf{m}) B_1^\dagger(\nabla) = \begin{bmatrix} 0 & -\frac{1}{\rho} D & \cdots & -\frac{1}{\rho} D & \cdots \\ -CD^T & 0 & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

$$N_1^{-\dagger}(\mathbf{m}) N_2^\dagger(\mathbf{m}) = \begin{bmatrix} 0 & 0 & \cdots & 0 & \cdots \\ 0 & 0 & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \omega_\ell y_\ell (C :: \Gamma) C^{-1} & 0 & \omega_\ell I_6 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

which should be compared with the explicit forward system (51). Similar to the state wavefield vector $\mathbf{w} = (\mathbf{v}^\dagger, \sigma^\dagger, \cdot, \xi_\ell^\dagger, \cdots)^\dagger$, the adjoint wavefield vector can also be denoted as $\bar{\mathbf{w}} = (\bar{\mathbf{v}}^\dagger, \bar{\sigma}^\dagger, \cdot, \bar{\xi}_\ell^\dagger, \cdots)^\dagger$ (the overbar indicates the adjoint). Let us denote the data residual

$$R_r \mathbf{w}(\mathbf{x},t) - \mathbf{d}(\mathbf{x}_r,t) = [\Delta d_{\mathbf{v}}^\dagger, \Delta d_\sigma^\dagger, \cdots, 0, \cdots]^\dagger. \tag{63}$$

As a consequence, the adjoint system (62) is expanded as

$$\rho \partial_t \bar{\mathbf{v}} = D\bar{\boldsymbol{\sigma}} + \sum_{\ell=1}^{L} D\bar{\boldsymbol{\xi}}_\ell + \Delta d_{\mathbf{v}} \tag{64a}$$

$$\partial_t \bar{\boldsymbol{\sigma}} = CD^T \bar{\mathbf{v}} + C\Delta d_\sigma \tag{64b}$$

$$\partial_t \bar{\boldsymbol{\xi}}_\ell - \omega_\ell \bar{\boldsymbol{\xi}}_\ell = \omega_\ell y_\ell (C :: \Gamma) C^{-1} \bar{\boldsymbol{\sigma}}, \ell = 1, \cdots, L. \tag{64c}$$

which enables efficient adjoint simulations through explicit time integration. This leads to the same solution as in implicit time integration system (60) which is more difficult to solve. As the final output, the gradient of the misfit functional should be the same, independent of the selected strategy as long as the consistent source terms are supplied. Adjoint sources defined in the adjoint equation (64) are related to the definition of the misfit function. Let us underline that the adjoint sources for stresses $\bar{\sigma}$ are linear combinations of different stress residuals weighted by matrix C rather than the data residual from a single component, as shown in equation (64b). When minimising the misfit function subjected to the condition of verifying the wave equation, one must notice that the adjoint equation (64) is not the same as the forward equation (46) due to the asymmetry induced by the coefficients of the memory variables. Also let us note that the memory variables are applied to the stresses in the forward simulation (46b), while the adjoint memory variables in the adjoint simulation are applied to the adjoint particle velocities in (64a). According to (64c), the memory variables for seismic attenuation is only related to the diagonal terms in the matrix Γ including all Q inverse.

This adjoint system of equations we have built based on standard minimisation under PDE-constraints is not the one promoted by Tarantola (1988); Charara et al. (2000) through integral expressions based on Green functions for which Fichtner & van Driel (2014) have provided explicitly the system of differential equations (labelled as equations (24), (27) and (28) in their article). This ad-hoc adjoint system turns out to be identical to the forward system of equations but the adjoint solutions are different from the definition of adjoint fields we obtain through the Lagrangian we have defined. In the appendix C, we shall exhibit the Lagrangian they have implicitly assumed and we shall show that this leads to equivalent gradient of the misfit function with respect to model parameters.

For a specific model perturbation δm, the variation of the objective function is expressed as

$$\delta\chi = \langle\frac{\partial\chi}{\partial\mathbf{m}},\delta\mathbf{m}\rangle_\Omega = \sum_{m\in\mathbf{m}}\langle\frac{\partial\chi}{\partial m},\delta m\rangle_\Omega = \langle\frac{\partial\chi}{\partial\rho},\delta\rho\rangle_\Omega + \sum_{I=1}^{6}\sum_{J=I}^{6}\langle\frac{\partial\chi}{\partial C_{IJ}},\delta C_{IJ}\rangle_\Omega + \sum_{I=1}^{6}\sum_{J=I}^{6}\langle\frac{\partial\chi}{\partial Q_{IJ}^{-1}},\delta Q_{IJ}^{-1}\rangle_\Omega \tag{65}$$

That is,

$$\begin{aligned}
&\langle\frac{\partial\chi}{\partial\mathbf{m}},\delta\mathbf{m}\rangle_\Omega = \langle\langle\bar{\mathbf{w}},\frac{\partial F(\mathbf{m},\mathbf{w})}{\partial\mathbf{m}}\rangle_T,\delta\mathbf{m}\rangle_\Omega = \sum_{m\in\mathbf{m}}\int_0^T \mathrm{d}t\int_\Omega \mathrm{d}\mathbf{x}\bar{\mathbf{w}}^\dagger\frac{\partial F(\mathbf{m},\mathbf{w})}{\partial m}\delta m\\
&=\sum_{m\in\mathbf{m}}\int_\Omega \mathrm{d}\mathbf{x}\int_0^T \mathrm{d}t\bar{\mathbf{w}}^\dagger\frac{\partial}{\partial m}\left(\begin{bmatrix}\rho I_3 & 0 & \cdots & 0 & \cdots\\ 0 & C^{-1} & \cdots & 0 & \cdots\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & \vdots & 0 & \frac{1}{\omega_\ell}I_6 & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots\end{bmatrix}\partial_t\mathbf{w} + \begin{bmatrix}0 & 0 & \cdots & 0 & \cdots\\ 0 & 0 & \cdots & y_\ell C^{-1}(C::\Gamma) & \cdots\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & \vdots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots\end{bmatrix}\mathbf{w} - \begin{bmatrix}\mathbf{f_v}\\ C^{-1}\mathbf{f}_\sigma\\ 0\\ \vdots\\ 0\end{bmatrix}\right)\delta m\\
&=\int_\Omega \mathrm{d}\mathbf{x}\delta\rho\left(\int_0^T \mathrm{d}t\bar{\mathbf{v}}^\dagger\partial_t\mathbf{v}\right) + \sum_{I=1}^{6}\sum_{J=I}^{6}\int_\Omega \mathrm{d}\mathbf{x}\delta C_{IJ}\left(\int_0^T \mathrm{d}t\bar{\boldsymbol{\sigma}}^\dagger\frac{\partial C^{-1}}{\partial C_{IJ}}(\partial_t\boldsymbol{\sigma}-\mathbf{f}_\sigma)\right)\\
&+\sum_{\ell=1}^{L}\sum_{I=1}^{6}\sum_{J=I}^{6}y_\ell\int_\Omega \mathrm{d}\mathbf{x}\delta Q_{IJ}^{-1}\left(\int_0^T \mathrm{d}t\bar{\boldsymbol{\sigma}}^\dagger C^{-1}(C::\frac{\partial\Gamma}{\partial Q_{IJ}^{-1}})\boldsymbol{\xi}_\ell\right),
\end{aligned}\tag{66}$$

where we have simple quantities

$$\left(\frac{\partial C}{\partial C_{IJ}}\right)_{ij} = \left(\frac{\partial\Gamma}{\partial Q_{IJ}^{-1}}\right)_{ij} = \begin{cases}1, & \text{if} \quad ij = IJ, JI\\ 0, & \text{otherwise.}\end{cases} \tag{67}$$

Invoking (46b) and the fact that

$$C^{-1}C = I \Rightarrow \frac{\partial C^{-1}}{\partial C_{IJ}}C + C^{-1}\frac{\partial C}{\partial C_{IJ}} = 0 \Rightarrow \frac{\partial C^{-1}}{\partial C_{IJ}} = -C^{-1}\frac{\partial C}{\partial C_{IJ}}C^{-1}, \tag{68}$$

we deduce the gradient of the misfit functional with respect to the density ρ, the stiffness constants C_{IJ}, and the quality factor Q_{IJ} given by the different components

$$\begin{aligned}
\frac{\partial\chi}{\partial\rho} &= \int_0^T \mathrm{d}t\bar{\mathbf{v}}^\dagger\partial_t\mathbf{v},\\
\frac{\partial\chi}{\partial C_{IJ}} &= -\int_0^T \mathrm{d}t\bar{\boldsymbol{\sigma}}^\dagger C^{-1}\frac{\partial C}{\partial C_{IJ}}C^{-1}(\partial_t\boldsymbol{\sigma}-\mathbf{f}_\sigma),\\
\frac{\partial\chi}{\partial Q_{IJ}^{-1}} &= \int_0^T \mathrm{d}t\bar{\boldsymbol{\sigma}}^\dagger C^{-1}(C::\frac{\partial\Gamma}{\partial Q_{IJ}^{-1}})(\sum_{\ell=1}^{L}y_\ell\boldsymbol{\xi}_\ell).
\end{aligned}\tag{69}$$

It is important to note that gradient components are simply zero-lag cross-correlations of quantities, independent of the number of relaxation mechanisms we consider. We end up with equivalent expressions as those proposed by Liu & Tromp (2006) and Vigh et al. (2014) in the non-dissipative case and the one for attenuation parameters by Fichtner & van Driel (2014). See the appendix C for extended discussion. Away from sources and receivers, we can express these gradients without stress through

$$\begin{aligned}
\frac{\partial\chi}{\partial\rho} &= \int_0^T \mathrm{d}t\bar{\mathbf{v}}^\dagger\partial_t\mathbf{v},\\
\frac{\partial\chi}{\partial C_{IJ}} &= -\int_0^T \mathrm{d}t\left(D^T\bar{\mathbf{u}}\right)^\dagger\frac{\partial C}{\partial C_{IJ}}\left(D^T\mathbf{v} - C^{-1}(C::\Gamma)\sum_{\ell=1}^{L}y_\ell\boldsymbol{\xi}_\ell\right),\\
\frac{\partial\chi}{\partial Q_{IJ}^{-1}} &= \int_0^T \mathrm{d}t\left(D^T\bar{\mathbf{u}}\right)^\dagger(C::\frac{\partial\Gamma}{\partial Q_{IJ}^{-1}})(\sum_{\ell=1}^{L}y_\ell\boldsymbol{\xi}_\ell),
\end{aligned}\tag{70}$$

where we have considered adjoint displacement $\bar{\mathbf{u}}$: a useful expression when stress is not directly available, as when considering the second-order hyperbolic system involving only particle velocities or displacements.

The computation of the gradient requires simultaneously accessing the forward wavefield $\mathbf{w}$ and the adjoint wavefield $\bar{\mathbf{w}}$, which has opposite time direction in the time-domain simulation. Some high level techniques might be useful such as optimal checkpointing strategy (Griewank & Walther 2000; Symes 2007; Anderson et al. 2012) and checkpointing-assisted reverse-forward simulation (CARFS) method (Yang et al. 2016) to access efficiently the incident wavefield when back propagating the adjoint wavefield. For the latter technique, the energy tracking during the reverse propagation of the incident field will help controlling the potential instability of this reverse propagation.

6 MODEL PARAMETRIZATION

In general, one may consider different families of parameters either for forward modeling or for parameter inversion. By following chain rules, we may deduce partial derivatives of the misfit function with respect to any set of parameters. The optimal choice of this set for FWI is still an open question (Innanen 2014; Korta et al. 2013) and may be case-dependent (Operto et al. 2013). In the following, we consider different sets of parameters and show how to relate each other through the chain rule.

6.1 Relating different attenuation parameters

For the modeling description, let us consider the stiffness matrix c_{ijkl} (or c_{IJ} using Voigt indexing) are parameterized by bulk modulus κ and shear moduli μ. Lamé parameters λ and μ, as well as the bulk modulus $\kappa = \lambda + \frac{2}{3}\mu$, are connected with compressional wave speed α and shear wave speed β through the following expression

$$c_{II} = \begin{cases} \kappa + \frac{4}{3}\mu = \lambda + 2\mu = \rho\alpha^2, & I \in \{1,2,3\}, \\ \mu = \rho\beta^2, & I \in \{4,5,6\}, \end{cases} \tag{71}$$

which are the diagonal terms of the stiffness matrix in isotropic case based on the Voigt indexing. In the presence of seismic attenuation, we may consider the quantities $Y_\ell^{II} c_{II}$, $I = 1, \cdots, 6$, using the parametrization with κ and μ

$$Y_\ell^{II} c_{II} = \begin{cases} \kappa Y_\ell^\kappa + \frac{4}{3}\mu Y_\ell^\mu, & I \in \{1,2,3\}; \\ \mu Y_\ell^\mu, & I \in \{4,5,6\} \end{cases} \tag{72}$$

or the parametrization with α and β:

$$Y_\ell^{II} c_{II} = \begin{cases} \rho\alpha^2 Y_\ell^\alpha, & I \in \{1,2,3\}; \\ \rho\beta^2 Y_\ell^\beta, & I \in \{4,5,6\} \end{cases} \tag{73}$$

Combining (71), (72) and (73) gives

$$Y_\ell^\kappa = \frac{\alpha^2 Y_\ell^\alpha - \frac{4}{3}\beta^2 Y_\ell^\beta}{\alpha^2 - \frac{4}{3}\beta^2}, Y_\ell^\mu = Y_\ell^\beta, \tag{74}$$

which relate the anelastic coefficients using different parametrization. The relation in (74) can be found in Kristek & Moczo (2003, eq. 4).

Similarly, according to the definition of Q, $\Im[M_{IJ}(\omega)] = Q_{IJ}^{-1}\Re[M_{IJ}(\omega)]$, with bulk and shear modulus parametrization we have

$$\Im[M_{II}(\omega)] = \begin{cases} \kappa Q_\kappa^{-1} + \frac{4}{3}\mu Q_\mu^{-1} & I \in \{1,2,3\} \\ \mu Q_\mu^{-1}, & I \in \{4,5,6\} \end{cases}, \tag{75}$$

or with P- and S- velocity parametrization

$$\Im[M_{II}(\omega)] = \begin{cases} \rho\alpha^2 Q_\alpha^{-1}, & I \in \{1,2,3\} \\ \rho\beta^2 Q_\beta^{-1}, & I \in \{4,5,6\} \end{cases}. \tag{76}$$

Combining (71), (75) and (76) gives

$$Q_\kappa^{-1} = \frac{\alpha^2 Q_\alpha^{-1} - \frac{4}{3}\beta^2 Q_\beta^{-1}}{\alpha^2 - \frac{4}{3}\beta^2}, Q_\mu^{-1} = Q_\beta^{-1}, \tag{77}$$

which can also be found in Stein & Wysession (2003, p. 192, eqs. 29 and 30) and Savage et al. (2010, eqs. 12 and 13). These expressions could be useful when changing the way to describe the attenuation.

6.2 3D isotropic viscoelastic inversion

As already underlined, the inversion may be carried out for different physical parameters. Some of the parameters have strong impacts on the data, while some of them might have small imprint. In 3D isotropic viscoelastic regime, we may be interested in inverting the bulk and shear moduli (or Lamé constants), or searching the solution for the P- and S- wave speeds, based on an objective function of the form

$$\chi'(\mathbf{m}') := \chi(\mathbf{m}), \tag{78}$$

where $\mathbf{m} = (\rho, C_{IJ}, Q_{IJ})^\dagger$, $\mathbf{m}' = (\rho, \alpha, \beta, Q_\alpha, Q_\beta)^\dagger$. A sound mathematical tool to shifting from one parameter group m to another m' is through the chain rule. Therefore, we end up with the gradient of the misfit function with respect to density, P- and S- wave as well as the corresponding Q in the following

$$\frac{\partial \chi'}{\partial \rho} = \frac{\partial \chi}{\partial \rho} + \sum_{I=1}^{6}\sum_{J=I}^{6} \frac{\partial \chi}{\partial C_{IJ}} \frac{\partial C_{IJ}}{\partial \rho}, \tag{79}$$

$$\frac{\partial \chi'}{\partial \alpha} = \sum_{I=1}^{6}\sum_{J=I}^{6} \frac{\partial \chi}{\partial C_{IJ}} \frac{\partial C_{IJ}}{\partial \alpha}, \frac{\partial \chi'}{\partial \beta} = \sum_{I=1}^{6}\sum_{J=I}^{6} \frac{\partial \chi}{\partial C_{IJ}} \frac{\partial C_{IJ}}{\partial \beta}, \tag{80}$$

$$\frac{\partial \chi'}{\partial Q_\alpha} = -Q_\alpha^{-2} \frac{\partial \chi'}{\partial Q_\alpha^{-1}}, \quad \frac{\partial \chi'}{\partial Q_\alpha^{-1}} = \sum_{I=1}^{6}\sum_{J=I}^{6} \frac{\partial \chi}{\partial Q_{IJ}^{-1}} \frac{\partial Q_{IJ}^{-1}}{\partial Q_\alpha^{-1}}, \tag{81}$$

$$\frac{\partial \chi'}{\partial Q_\beta} = -Q_\beta^{-2} \frac{\partial \chi'}{\partial Q_\beta^{-1}}, \quad \frac{\partial \chi'}{\partial Q_\beta^{-1}} = \sum_{I=1}^{6}\sum_{J=I}^{6} \frac{\partial \chi}{\partial Q_{IJ}^{-1}} \frac{\partial Q_{IJ}^{-1}}{\partial Q_\beta^{-1}}. \tag{82}$$

One may rewrite the stiffness matrix C as

$$C = \begin{bmatrix} \lambda+2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda+2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda+2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{bmatrix} = \begin{bmatrix} \rho\alpha^2 & \rho(\alpha^2-2\beta^2) & \rho(\alpha^2-2\beta^2) & 0 & 0 & 0 \\ \rho(\alpha^2-2\beta^2) & \rho\alpha^2 & \rho(\alpha^2-2\beta^2) & 0 & 0 & 0 \\ \rho(\alpha^2-2\beta^2) & \rho(\alpha^2-2\beta^2) & \rho\alpha^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho\beta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho\beta^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho\beta^2 \end{bmatrix}. \tag{83}$$

As a consequence, we have

$$\frac{\partial C}{\partial \rho} = \begin{bmatrix} \alpha^2 & \alpha^2-2\beta^2 & \alpha^2-2\beta^2 & 0 & 0 & 0 \\ \alpha^2-2\beta^2 & \alpha^2 & \alpha^2-2\beta^2 & 0 & 0 & 0 \\ \alpha^2-2\beta^2 & \alpha^2-2\beta^2 & \alpha^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta^2 \end{bmatrix}, \tag{84}$$

$$\frac{\partial C}{\partial \alpha} = 2\rho\alpha \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \frac{\partial C}{\partial \beta} = 2\rho\beta \begin{bmatrix} 0 & -2 & -2 & 0 & 0 & 0 \\ -2 & 0 & -2 & 0 & 0 & 0 \\ -2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{85}$$

The matrix including all Q inverse becomes

$$\Gamma = \begin{bmatrix} Q_\alpha^{-1} & Q_\alpha^{-1}-2Q_\beta^{-1} & Q_\alpha^{-1}-2Q_\beta^{-1} & 0 & 0 & 0 \\ Q_\alpha^{-1}-2Q_\beta^{-1} & Q_\alpha^{-1} & Q_\alpha^{-1}-2Q_\beta^{-1} & 0 & 0 & 0 \\ Q_\alpha^{-1}-2Q_\beta^{-1} & Q_\alpha^{-1}-2Q_\beta^{-1} & Q_\alpha^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & Q_\beta^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & Q_\beta^{-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & Q_\beta^{-1} \end{bmatrix}, \tag{86}$$

yielding

$$\frac{\partial \Gamma}{\partial Q_\alpha^{-1}} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \frac{\partial \Gamma}{\partial Q_\beta^{-1}} = \begin{bmatrix} 0 & -2 & -2 & 0 & 0 & 0 \\ -2 & 0 & -2 & 0 & 0 & 0 \\ -2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{87}$$

Thus, all partial derivatives in (79), i.e. $\partial C_{IJ}/\partial \rho$, $\partial C_{IJ}/\partial \alpha$,$\partial C_{IJ}/\partial \beta$, $\partial Q_{IJ}^{-1}/\partial Q_\alpha^{-1}$, $\partial Q_{IJ}^{-1}/\partial Q_\beta^{-1}$, are now available explicitly in equations (84) and (87).

7 CONCLUSION

In this paper, we formulate in a consistent way the 3D multiparameter FWI in viscoelastic media based on GMB using arbitrary number of attenuation mechanisms. According to the elastic-viscoelastic correspondence principle in the Fourier domain, we have developed an energy analysis to determine the stable energy attenuating conditions of viscoelastic system. We have reformulated the least-squares optimization problem for estimating the anelastic coefficients by introducing a new set of constants, which linearly relates the Q parameter and the anelastic coefficients. Equipped with these constants, the Q parameter can be explicitly displaced into the viscoelastic system to be further inverted in FWI framework. By introducing the standard Lagrangian multipliers into the matrix expression of the first-order wave equation with implicit time integration, we have built a systematic formulation of multiparameter FWI for the most general anisotropic viscoelastic wave equation, while the equivalent form of the state and adjoint equation with explicit time integration is available to be resolved efficiently. The adjoint and forward systems are different but one can manipulate the Lagrangian expression to propose an adjoint system similar to the forward system, thanks to the linearity of the wave equation and the properties of the Green's functions. In the 3D isotropic viscoelastic settings, the anelastic coefficients and the quality factors using bulk and shear moduli parametrization can be related to the counterparts using P- and S- velocity parametrization. The final gradient with a specific parametrization of model parameters can be found from the gradient of the misfit function with another parametrization through the chain rule. These mathematical development should facilitate the application of FWI including a frequency-independent Q inversion in seismology and exploration geophysics in a consistent framework. Numerical examples should be performed to disclose its potential, and will be strongly dependent on the geometry of the target and the acquisition aside the medium properties.

ACKNOWLEDGMENTS

We thank the communications with Wei Zhou, P.T. Trinh and Stéphane Operto, and Bruno Lombard for initial discussion about the energy balance for Maxwell bodies. We appreciate very much for the inspiring comments from René-Édouard Plessix on the total energy of the viscoelastic system, and the valuable suggestions from Thomas Bohlen to improve the manuscript. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO, SCHLUMBERGER, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (*https://ciment.ujf-grenoble.fr*), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS under the allocation 046091 made by GENCI.

References

Alfrey, T., 1948. *Mechanical property of high polymer*, John Wiley & Sons, New York.

Anderson, J. E., Tan, L., & Wang, D., 2012. Time-reversal checkpointing methods for RTM and FWI, *Geophysics*, **77**, S93–S103.

Auld, B. A., 1990. *Acoustic fields and waves in solids, volume 1 and 2*, Krieger Publishing Co.

Azimi, S. A., Kalinin, A., Kalinin, V., & Pivovarov, B., 1968. Impulse and transient characteristics of media with linear and quadratic absorption laws, *Izv. Phys. Solid Earth*, **2**, 88–93.

Bai, J., Yingst, D., Bloor, R., & Leveille, J., 2014. Viscoacoustic waveform inversion of velocity structures in the time domain, *Geophysics*, **79**(3), R103–R119.

Bécache, E., Ezziani, A., & Joly, P., 2005. A mixed finite element approach for viscoelastic wave propagation, *Computational Geosciences*, **8**(3), 255–299.

Ben-Menahem, A. & Singh, S., 1981. *Seismic waves and sources*, Springer-Verlag, New York :.

Blanch, J., Robertson, J. O. A., & Symes, W. W., 1995. Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique, *Geophysics*, **60**, 176–184.

Bohlen, T., 2002. Parallel 3-D viscoelastic finite-difference seismic modeling, *Computers & Geosciences*, **28**, 887–899.

Brossier, R., Operto, S., & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, **74**(6), WCC105–WCC118.

Brull, M., 1953. A structural theory incorporating the effect of time-dependent elasticity, in *Proc. 1st Midwestern Conf. Solid Mech*, vol. 141.

Caputo, M., 1967. Linear models of dissipation whose q is almost frequency independent–ii, *Geophysical Journal International*, **13**(5), 529–539.

Carcione, J. & Cavallini, F., 1993. Energy balance and fundamental relations in anisotropic-viscoelastic media, *Wave motion*, **18**(1), 11–20.

Carcione, J. M., 2001. *Wave fields in real media, Theory and numerical simulation of wave propagation in anisotropic, an elastic and porous media*, Pergamon Press.

Carcione, J. M., Kosloff, D., & Kosloff, R., 1988a. Wave-propagation simulation in an elastic anisotropic (transversely isotropic) solid, *Quarterly Journal of Mechanics and Applied Mathematics*, **41**(3), 319–345.

Carcione, J. M., Kosloff, D., & Kosloff, R., 1988b. Wave propagation simulation in a linear viscoacoustic medium, *Geophysical Journal International*, **93**(2), 393–401.

Carcione, J. M., Kosloff, D., & Kosloff, R., 1988c. Wave propagation simulation in a linear viscoelastic medium, *Geophysical Journal International*, **95**(3), 597–611.

Casula, G. & Carcione, J., 1992. Generalized mechanical model analogies of linear viscoelastic behaviour, *Bollettino di geofisica teorica ed applicata*, **XXXIV**(136).

Červený, V. & Pšenčík, I., 2006. Energy flux in viscoelastic anisotropic media, *Geophysical Journal International*, **166**(3), 1299–1317.

Charara, M., Barnes, C., & Tarantola, A., 2000. Full waveform inversion of seismic data for a viscoelastic medium, in *Methods and Applications of Inversion*, pp. 68–81, Springer.

Cheng, X., Jiao, K., Sun, D., Vigh, D., et al., 2015. A new approach of visco-acoustic waveform inversion in the time domain, in *2015 SEG Annual Meeting*, Society of Exploration Geophysicists.

Christensen, R., 1982. *Theory of viscoelasticity: an introduction*, Elsevier.

Emmerich, H. & Korn, M., 1987. Incorporation of attenuation into time-domain computation of seismic wavefield, *Geophysics*, **52**, 1252–1264.

Ferry, J. D., 1961. *Viscoelastic properties of polymers*, John Wiley & Sons, 3rd edn.

Fichtner, A. & van Driel, M., 2014. Models and Fréchet kernels for frequency-(in)dependent Q, *Geophysical Journal International*, **198(3)**, 1878–1889.

Futterman, W., 1962. Dispersive body waves, *Journal Geophysical Research*, **67**, 5279–5291.

Gauthier, O., Virieux, J., & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics*, **51**(7), 1387–1403.

Gholami, Y., Brossier, R., Operto, S., Prieux, V., Ribodetti, A., & Virieux, J., 2013. Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 2: application to Valhall, *Geophysics*, **78**(2), R107–R124.

Gokhberg, A. & Fichtner, A., 2015. Full-waveform inversion on heterogeneous hpc systems, *Computers & Geosciences*.

Griewank, A. & Walther, A., 2000. Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation, *ACM Transactions on Mathematical Software (TOMS)*, **26**(1), 19–45.

Hicks, G. J. & Pratt, R. G., 2001. Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit, *Geophysics*, **66**(2), 598–612.

Innanen, K. A., 2014. Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion, *Geophysical Journal International*, **199**(2), 717–734.

Kamei, R. & Pratt, R. G., 2013. Inversion strategies for visco-acoustic waveform inversion, *Geophysical Journal International*, **194**, 859–894.

Kjartansson, E., 1979. Constant Q-wave propagation and attenuation, *J. geophys. Res*, **84**(B9), 4737–4748.

Kolsky, H., 1956. The propagation of stress pulses in viscoelastic solids, *Philosophical Magazine*, **1**, 693–710.

Korta, N., Fichtner, A., & Sallares, V., 2013. Block-diagonal approximate hessian for preconditioning in full waveform inversion, in *Expanded Abstracts, 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013, London*.

Kristek, J. & Moczo, P., 2003. Seismic-wave propagation in viscoelastic media with material discontinuities: A 3d fourth-order staggered-grid finite-difference modeling, *Bulletin of the Seismological Society of America*, **93**(5), 2273–2280.

Kurzmann, A., Przebindowska, A., Kohn, D., & Bohlen, T., 2013. Acoustic full waveform tomography in the presence of attenuation: a sensitivity analysis, *Geophysical Journal International*, **195(2)**, 985–1000.

Lailly, P., 1983. The seismic problem as a sequence of before-stack migrations, in *Conference on Inverse Scattering: Theory and Applications*, SIAM, Philadelphia.

Lee, E., 1960. *Stress analysis for viscoelastic bodies*, Academic Press New York.

Liu, H.-P., Anderson, D. L., & Kanamori, H., 1976. Velocity dispersion due to anelasticity; implications for seismology and mantle composition, *Geophysical Journal International*, **47**(1), 41–58.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bulletin of the Seismological Society of America*, **96**(6), 2383–2397.

Moczo, P. & Kristek, J., 2005. On the rheological models used for time-domain methods of seismic wave propagation, *Geophysical Research Letters*, **32**(1).

Moczo, P., Kristek, J., Galis, M., Pazak, P., & Balazovjech, M., 2007a. The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion, *Acta Physica Slovaca*, **52**(2), 177–406.

Moczo, P., Robertsson, J., & Eisner, L., 2007b. The finite-difference time-domain method for modeling of seismic wave propagation, in *Advances in Wave Propagation in Heterogeneous Earth*, Advances in Geophysics, pp. 421–516, eds Wu, R.-S., Maupin, V., & Dmowska, R., Elsevier/Academic Press, San Diego.

Mora, P. R., 1987. Nonlinear two-dimensional elastic inversion of multi-offset seismic data, *Geophysics*, **52**, 1211–1228.

Operto, S., Virieux, J., Ribodetti, A., & Anderson, J. E., 2009. Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media, *Geophysics*, **74 (5)**, T75–T95.

Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A., & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice, *The Leading Edge*, **Special section Full Waveform Inversion**(September), 1040–1054.

Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., & Virieux, J., 2015. Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *Geophysical Journal International*, **202**(2), 1362–1391.

Pipkin, A. C., 1986. *Lectures on viscoelasticity theory*, vol. 7, Springer Science & Business Media.

Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503.

Pratt, R. G., Shin, C., & Hicks, G. J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophysical Journal International*, **133**, 341–362.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O., & Kommedal, J., 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophysical Journal International*, **187**, 1495–1515.

Prieux, V., Brossier, R., Operto, S., & Virieux, J., 2013a. Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 1: imaging compressional wavespeed, density and attenuation, *Geophysical Journal International*, **194**(3), 1640–1664.

Prieux, V., Brossier, R., Operto, S., & Virieux, J., 2013b. Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 2: imaging compressional and shear-wave velocities, *Geophysical Journal International*, **194**(3), 1665–1681.

Read Jr, W., 1950. Stress analysis for compressible viscoelastic materials, *Journal of Applied Physics*, **21**(7), 671–674.

Romanowicz, B., 1995. A global tomographic model of shear attenuation in the upper mantle, *Journal of Geophysical Research*, **100**, 12375–12394.

Romanowicz, B. & Mitchell, B., 2007. Deep Earth structure - Q of the Earth from crust to core, in *Treatise of Geophysics, volume 1: Seismology and structure of the Earth*, pp. 731–774, eds Romanowicz, B. & Diewonski, A., Elsevier.

Savage, B., Komatitsch, D., & Tromp, J., 2010. Effects of 3d attenuation on seismic wave amplitude and phase measurements, *Bulletin of the Seismological Society of America*, **100**(3), 1241–1251.

Scott, N., 1997. Energy flux and dissipation of inhomogeneous plane wave forms in kelvin-voigt viscoelasticity, *International journal of engineering science*, **35**(6), 561–572.

Shin, C. & Cha, Y. H., 2008. Waveform inversion in the Laplace domain, *Geophysical Journal International*, **173**(3), 922–931.

Shin, C. & Cha, Y. H., 2009. Waveform inversion in the Laplace-Fourier domain, *Geophysical Journal International*, **177**, 1067–1079.

Shin, J., Ha, W., Jun, H., Min, D.-J., & Shin, C., 2014. 3D Laplace-domain full waveform inversion using a single GPU card, *Computers & Geosciences*, **67**, 1–13.

Sips, R., 1951. General theory of deformation of viscoelastic substances, *Journal of Polymer Science*, **7**(2-3), 191–205.

Sirgue, L., 2006. The importance of low frequency and large offset in waveform inversion, in *68th EAGE Conference & Exhibition, Vienna, EAGE*, p. A037.

Sirgue, L. & Pratt, R. G., 2004. Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies, *Geophysics*, **69**(1), 231–248.

Stein, S. & Wysession, M., 2003. *An Introduction to Seismology, Earthquakes and Earth Structure*, Blackwell Publishing.

Strick, E., 1967. The determination of q, dynamic viscosity and transient creep curves from wave propagation measurements, *Geophysical Journal Interna-*

tional, **13**(1-3), 197–218.
Symes, W. W., 2007. Reverse time migration with optimal checkpointing, *Geophysics*, **72**(5), SM213–SM221.
Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
Tarantola, A., 1987. *Inverse problem theory: methods for data fitting and model parameter estimation*, Elsevier, New York.
Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms including elasticity and attenuation, *Pure and Applied Geophysics*, **128**, 365–399.
Tromp, J., Tape, C., & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophysical Journal International*, **160**, 195–216.
Ursin, B. & Toverud, T., 2002. Comparison of seismic dispersion and attenuation models, *Studia geophysica et geodaetica*, **46**(2), 293–320.
Vigh, D. & Starr, E. W., 2008. 3D prestack plane-wave, full-waveform inversion, *Geophysics*, **73**(5), VE135–VE144.
Vigh, D., Jiao, K., Watts, D., & Sun, D., 2014. Elastic full-waveform inversion application using multicomponent measurements of seismic data collection, *Geophysics*, **79**(2), R63–R77.
Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
Yang, P., Gao, J., & Wang, B., 2015. A graphics processing unit implementation of time-domain full-waveform inversion, *Geophysics*, **80**(3), F31–F39.
Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016. Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method, *Geophysics, Accepted*.
Zhou, W., Brossier, R., Operto, S., & Virieux, J., 2015. Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation, *Geophysical Journal International*, **202**(3), 1535–1554.

APPENDIX A: GZB/GMB-BASED VISCOELASTIC WAVE EQUATION IN 1D

The rheology for a 1D attenuating medim can be described through the linear relation (Christensen 1982)

$$\sigma = \psi *_t \dot{\epsilon} = \dot{\psi} *_t \epsilon, \tag{A.1}$$

where the stress is denoted by the symbol σ, the deformation by ϵ. The dot over a variable stands for the time derivative. Following Casula & Carcione (1992), the relaxation function ψ for the generalized Zener model or standard linear solid (SLS) is defined by the expression

$$\psi(t) = M_r \left(1 - \frac{1}{L}\sum_{\ell=1}^{L}(1 - \frac{\tau_{\epsilon\ell}}{\tau_{\sigma\ell}})e^{-\frac{t}{\tau_{\sigma\ell}}}\right) H(t), \tag{A.2}$$

where the relaxed modulus is denoted by M_r, two characteristic relaxation times by $\tau_{\sigma\ell}$ and $\tau_{\epsilon\ell}$, with a number L of standard linear solids. We must consider that the relaxation time $\tau_{\sigma\ell}$ is always higher than the relaxation time $\tau_{\epsilon\ell}$ because of the energy dissipation. The Heaviside function is denoted by $H(t)$. Let us define

$$\tilde{Y}_\ell = \frac{1}{L}(\frac{\tau_{\epsilon\ell}}{\tau_{\sigma\ell}} - 1),\ \ell = 1, \ldots, L, \tag{A.3}$$

and the circular frequency $\omega_\ell = 1/\tau_{\sigma\ell}$. The equation (A.2) can be rewritten as

$$\psi(t) = M_r \left(1 + \sum_{\ell=1}^{L}\tilde{Y}_\ell e^{-\omega_\ell t}\right) H(t). \tag{A.4}$$

Let us note that $\tilde{Y}_\ell$ in (A.3) is exactly the τ of the well-known τ-method defined by Blanch et al. (1995, equation 13) with only $L = 1$ mechanism. The unrelaxed modulus M_u is defined at the origin time through the expression

$$M_u = \lim_{t\to 0^+}\psi(t) = M_r(1 + \sum_{\ell=1}^{L}\tilde{Y}_\ell). \tag{A.5}$$

We introduce the non-dimensional anelastic parameter

$$Y_\ell = \frac{M_r}{M_u}\tilde{Y}_\ell, \ell = 1, \ldots, L. \tag{A.6}$$

The relation in (A.4) and (A.5) becomes

$$M_u = M_r(1 + \sum_{\ell=1}^{L}\tilde{Y}_\ell) = M_r + M_u\sum_{\ell=1}^{L}Y_\ell \Rightarrow M_r = M_u(1 - \sum_{\ell=1}^{L}Y_\ell). \tag{A.7}$$

and

$$\psi(t) = M_r \left(1 + \sum_{\ell=1}^{L}\tilde{Y}_\ell e^{-\omega_\ell t}\right) H(t) = M_u(1 - \sum_{\ell=1}^{L}Y_\ell(1 - e^{-\omega_\ell t}))H(t). \tag{A.8}$$

We may consider as well that dissipation effects are always more important for compressive waves than for shear waves inducing higher values of the coefficient Y_ℓ for compression mode than for shear mode. The first time derivative of the equation (A.8) gives

$$\begin{aligned}\dot{\psi}(t) &= M_u(1-\sum_{\ell=1}^{L} Y_\ell(1-e^{-\omega_\ell t}))\delta(t) + M_u\partial_t[1-\sum_{\ell=1}^{L} Y_\ell(1-e^{-\omega_\ell t})]H(t)\\ &= M_u\delta(t) - M_u\sum_{\ell=1}^{L} Y_\ell\omega_\ell e^{-\omega_\ell t}H(t).\end{aligned} \tag{A.9}$$

Thus, the rheological relation becomes

$$\begin{aligned}\sigma =&\dot{\psi} *_t \epsilon\\ =&\left(M_u\delta(t) - M_u\sum_{\ell=1}^{L} Y_\ell\omega_\ell e^{-\omega_\ell t}H(t)\right) *_t \epsilon(t)\\ =&M_u\epsilon(t) - M_u\sum_{\ell=1}^{L} Y_\ell\omega_\ell e^{-\omega_\ell t}H(t) *_t \epsilon(t)\\ =&M_u\epsilon(t) - M_u\sum_{\ell=1}^{L} Y_\ell\zeta(t),\end{aligned} \tag{A.10}$$

where we have introduced new variables $\zeta_\ell(t) := \omega_\ell e^{-\omega_\ell t}H(t) *_t \epsilon(t)$. The time derivative of these intermediate variables is

$$\begin{aligned}\dot{\zeta}_\ell(t) &= \partial_t[\omega_\ell e^{-\omega_\ell t}H(t)] *_t \epsilon(t)\\ &= [-\omega_\ell^2 e^{-\omega_\ell t}H(t) + \omega_\ell e^{-\omega_\ell t}\delta(t)] *_t \epsilon(t)\\ &= -\omega_\ell^2 e^{-\omega_\ell t}H(t) * \epsilon(t) + \omega_\ell\delta(t) *_t \epsilon(t)\\ &= \omega_\ell[-\zeta_\ell(t) + \epsilon(t)],\ \ell = 1,\ldots,L.\end{aligned} \tag{A.11}$$

Let us introduce the so-called auxiliary variables $\xi_\ell(t)$ as the time derivatives of ζ_ℓ for recovering the deformation derivative and, therefore, particle velocities. Combining the equations (A.10) and (A.11) gives

$$\begin{cases}\dot{\sigma} = M_u\dot{\epsilon} - M_u\sum_{\ell=1}^{L} Y_\ell\xi_\ell\\ \dot{\xi}_\ell = \omega_\ell[-\xi_\ell + \dot{\epsilon}],\ \ell = 1,\ldots,L.\end{cases} \tag{A.12}$$

The above construction was established by Emmerich & Korn (1987, their equations 12 and 13) for generalized Maxwell body (GMB) which is shown to be equivalent to generalized Zener body (GZB) (Moczo & Kristek 2005; Moczo et al. 2007a,b).

Due to the symmetry of the stress tensor, in the isotropic assumption the Hooke's law reads

$$\sigma_{ij} = \lambda\epsilon_{kk}\delta_{ij} + 2\mu\epsilon_{ij} = \kappa\epsilon_{kk}\delta_{ij} + 2\mu\left(\epsilon_{ij} - \frac{1}{3}\epsilon_{kk}\delta_{ij}\right), \tag{A.13}$$

where λ and μ are Lamé parameters related to the bulk modulus $\kappa = \lambda + \frac{2}{3}\mu$. In order to incorporate seismic attenuation based on the physically meaningful mechanism, Kristek & Moczo (2003); Moczo & Kristek (2005) generalized the elastic Hooke's law through the relation

$$\sigma_{ij} = \kappa\epsilon_{kk}\delta_{ij} + 2\mu\left(\epsilon_{ij} - \frac{1}{3}\epsilon_{kk}\delta_{ij}\right) - \sum_{\ell=1}^{L}\left[\kappa Y_\ell^\kappa\zeta_\ell^{kk}\delta_{ij} + 2\mu Y_\ell^\mu\left(\zeta_\ell^{ij} - \frac{1}{3}\zeta_\ell^{kk}\delta_{ij}\right)\right]. \tag{A.14}$$

Note that the corresponding anelastic coefficients Y_ℓ^κ and Y_ℓ^μ for unrelaxed bulk and shear moduli have been introduced in the isotropic generalization (A.14). The new quantities, called memory variables ζ_ℓ^{ij}, satisfy $\dot{\zeta}_\ell^{ij} + \omega_\ell\zeta_\ell^{ij} = \omega_\ell\epsilon_{ij}, \ell = 1,\ldots,L$, where a series of L mechanisms are used to characterize the attenuation rheology. Applying the first time derivative to equations (A.14), leading to isotropic viscoelastic wave equation

$$\begin{cases}\rho\dot{v}_i = \partial_j\sigma_{ij}\\ \dot{\sigma}_{ij} = \kappa\dot{\epsilon}_{kk}\delta_{ij} + 2\mu\left(\dot{\epsilon}_{ij} - \frac{1}{3}\dot{\epsilon}_{kk}\delta_{ij}\right) - \sum_{\ell=1}^{L}\left[\kappa Y_\ell^\kappa\xi_\ell^{kk}\delta_{ij} + 2\mu Y_\ell^\mu\left(\xi_\ell^{ij} - \frac{1}{3}\xi_\ell^{kk}\delta_{ij}\right)\right],\\ \dot{\xi}_\ell^{ij} + \omega_\ell\xi_\ell^{ij} = \omega_\ell\dot{\epsilon}_{ij}, \ell = 1,\ldots,L,\end{cases} \tag{A.15}$$

where we introduce another set of memory variables ξ_ℓ^{ij} to denote the time derivatives of ζ_ℓ^{ij} expressed as $\xi_\ell^{ij} = \dot{\zeta}_\ell^{ij}$ for the numerical implementation of the equation (A.15).

APPENDIX B: ENERGY ANALYSIS IN TIME DOMAIN

We have performed the energy analysis in the main core of this article for the general anisotropic case. Here we would like to illustrate that it can be obtained directly for the isotropic case without considering the convolution procedure we have used.

B1 Elastic case

For the energy definition, let us first start with the elastic case where the following total energy $E = E_k + E_p$ is decomposed into a kinematic energy denoted by E_k and a strain energy denoted by E_p through

$$\begin{cases} E_k = \frac{1}{2}\langle \rho\mathbf{v}, \mathbf{v}\rangle_\Omega, \\ E_p = \frac{1}{2}\sum_i\sum_j\langle \sigma_{ij}, \epsilon_{ij}\rangle_\Omega = \frac{1}{2}\sum_i\sum_j\sum_k\sum_l\langle c_{ijkl}\epsilon_{kl}, \epsilon_{ij}\rangle_\Omega. \end{cases} \tag{B.1}$$

The energy E_p is also called potential energy, implying that the work does not depends on the path. As can be seen above, the kinetic energy is related to the particle velocity $\mathbf{v}$–the time derivative of the displacement, while the potential energy is related to the spatial derivatives of the displacement ϵ_{ij}. Both E_k and E_p are nonnegative quantities assuming a positive definite 4-th order tensor c_{ijkl}. For the energy preservation in elastodynamics, we need to show the following identity

$$\partial_t E = \langle \rho\dot{\mathbf{v}}, \mathbf{v}\rangle_\Omega + \sum_i\sum_j\sum_k\sum_l\langle c_{ijkl}\epsilon_{kl}, \dot{\epsilon_{ij}}\rangle_\Omega = 0, \tag{B.2}$$

thanks to the linear property of the Hooke law.

The first time derivative of the potential energy is

$$\begin{aligned} \partial_t E_p =& \sum_i\sum_j\sum_k\sum_l\langle c_{ijkl}\epsilon_{kl}, \dot{\epsilon}_{ij}\rangle_\Omega \\ =& \sum_i\sum_j\langle \sigma_{ij}, \dot{\epsilon}_{ij}\rangle_\Omega = \frac{1}{2}\sum_i\sum_j\langle \sigma_{ij}, v_{i,j} + v_{j,i}\rangle_\Omega \\ =& \frac{1}{2}\sum_i\sum_j\langle \sigma_{ij}, v_{i,j}\rangle_\Omega + \frac{1}{2}\sum_i\sum_j\langle \sigma_{ij}, v_{j,i}\rangle_\Omega \\ =& \frac{1}{2}\sum_i\sum_j\langle \sigma_{ij}, v_{i,j}\rangle_\Omega + \frac{1}{2}\sum_i\sum_j\langle \sigma_{ji}, v_{j,i}\rangle_\Omega \\ =& \sum_i\sum_j\langle \sigma_{ij}, v_{i,j}\rangle_\Omega, \end{aligned} \tag{B.3}$$

using the symmetry of the stress tensor. The Newton's law

$$\rho\dot{v}_i = \partial_j\sigma_{ij} = \sum_j \sigma_{ij,j} \tag{B.4}$$

can be used in the inner product

$$\langle \rho\dot{v}_i, v_i\rangle_\Omega = \sum_j\langle v_i, \sigma_{ij,j}\rangle_\Omega = -\sum_j\langle v_{i,j}, \sigma_{ij}\rangle_\Omega. \tag{B.5}$$

From (B.2), we may deduce the following time evolution of the kinetic energy

$$\partial_t E_k = \sum_i\langle \rho\dot{v}_i, v_i\rangle_\Omega = -\sum_i\sum_j\langle v_{i,j}, \sigma_{ij}\rangle_\Omega. \tag{B.6}$$

Summing over (B.3) and (B.6) concludes the validity of the equation (B.2). Therefore, the elastic system is energy preserving with varying time.

B2 Isotropic viscoelastic case

The isotropic case could be deduced from the anisotropic one: we detail here a direct analysis of the energy balance for this particular case. The non-negative total energy is defined as

$$\begin{aligned} E =& E_k + E_s, \\ E_k =& \frac{1}{2}\langle \rho\mathbf{v}, \mathbf{v}\rangle_\Omega, \\ E_s =& \frac{1}{2}\langle (M_r^\kappa - \frac{2}{3}M_r^\mu)\epsilon_{kk}, \epsilon_{kk}\rangle_\Omega + \frac{1}{2}\sum_i\sum_j\langle 2M_r^\mu\epsilon_{ij}, \epsilon_{ij}\rangle_\Omega \\ &+ \frac{1}{2}\sum_{\ell=1}^{L}\sum_i\sum_j\langle \frac{2\mu Y_\ell^\mu}{\omega_\ell^2}\xi_\ell^{ij}, \xi_\ell^{ij}\rangle_\Omega + \frac{1}{2}\sum_{\ell=1}^{L}\langle \frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell^2}\xi_\ell^{kk}, \xi_\ell^{kk}\rangle_\Omega, \end{aligned} \tag{B.7}$$

where the quantity $\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu$ is assumed to be non-negative; $M_r^\kappa = \kappa(1 - \sum_{\ell=1}^L Y_\ell^\kappa)$ and $M_r^\mu = \mu(1 - \sum_{\ell=1}^L Y_\ell^\mu)$ are the relaxed bulk modulus and the relaxed shear modulus, respectively. Note that the strain energy E_s does not amount to potential energy in viscoelastics because the work performed from one point to the other one now depends on the path between them (this is the reason why the system has memory about the past history).

For the memory variables, we have following ordinary differential equations

$$\partial_t \xi_\ell^{ij} + \omega_\ell \xi_\ell^{ij} = \omega_\ell \dot{\epsilon}_{ij} \Rightarrow \xi_\ell^{ij} = -\frac{1}{\omega_\ell}\partial_t \xi_\ell^{ij} + \omega_\ell \dot{\epsilon}_{ij}, \tag{B.8}$$

which can be written as

$$\frac{1}{\omega_\ell}\partial_t \xi_\ell^{kk} = -\xi_\ell^{kk} + \dot{\epsilon}_{kk}. \tag{B.9}$$

Taking the inner product over Ω with $\frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij}$ and $\frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{kk}$ gives

$$\langle \frac{2\mu Y_\ell^\mu}{\omega_\ell^2}\partial_t \xi_\ell^{ij}, \xi_\ell^{ij}\rangle_\Omega = -\langle \frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij}, \xi_\ell^{ij}\rangle_\Omega + \langle \dot{\epsilon}_{ij}, \frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij}\rangle_\Omega \tag{B.10}$$

and

$$\langle \frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell^2}\partial_t \xi_\ell^{kk}, \xi_\ell^{kk}\rangle_\Omega = -\langle \frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{kk}, \xi_\ell^{kk}\rangle_\Omega + \langle \dot{\epsilon}_{kk}, \frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{kk}\rangle_\Omega. \tag{B.11}$$

Therefore, we have the following equalities

$$\sum_{\ell=1}^{L}\sum_i\sum_j \langle \frac{2\mu Y_\ell^\mu}{\omega_\ell^2}\partial_t \xi_\ell^{ij}, \xi_\ell^{ij}\rangle_\Omega = -\sum_{\ell=1}^{L}\sum_i\sum_j \langle \frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij}, \xi_\ell^{ij}\rangle_\Omega + \sum_{\ell=1}^{L}\sum_i\sum_j \langle \dot{\epsilon}_{ij}, \frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij}\rangle_\Omega, \tag{B.12}$$

$$\sum_{\ell=1}^{L}\langle \frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell^2}\partial_t \xi_\ell^{kk}, \xi_\ell^{kk}\rangle_\Omega = -\sum_{\ell=1}^{L}\langle \frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{kk}, \xi_\ell^{kk}\rangle_\Omega + \sum_{\ell=1}^{L}\langle \dot{\epsilon}_{kk}, \frac{\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{kk}\rangle_\Omega. \tag{B.13}$$

According to equations (A.13) and (B.8), setting $i = j$ gives

$$\begin{aligned}
\sigma_{ii} =& (\kappa - \frac{2}{3}\mu)\epsilon_{kk} + 2\mu\epsilon_{ii} - \sum_{\ell=1}^{L}[(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)\zeta_\ell^{kk} + 2\mu Y_\ell^\mu \zeta_\ell^{ii}] \\
=& (\kappa - \frac{2}{3}\mu)\epsilon_{kk} + 2\mu\epsilon_{ii} - \sum_{\ell=1}^{L}[(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)(-\frac{1}{\omega_\ell}\dot{\zeta}_\ell^{kk} + \epsilon_{kk}) + 2\mu Y_\ell^\mu(-\frac{1}{\omega_\ell}\dot{\zeta}_\ell^{kk} + \epsilon_{ii})] \\
=& (\kappa(1 - \sum_{\ell=1}^{L} Y_\ell^\kappa) - \frac{2}{3}\mu(1 - \sum_{\ell=1}^{L} Y_\ell^\mu))\epsilon_{kk} + 2\mu(1 - \sum_{\ell=1}^{L} Y_\ell^\mu)\epsilon_{ii} + \frac{1}{\omega_\ell}\sum_{\ell=1}^{L}[(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)\dot{\zeta}_\ell^{kk} + 2\mu Y_\ell^\mu \dot{\zeta}_\ell^{ii}] \\
=& (M_r^\kappa - \frac{2}{3}M_r^\mu)\epsilon_{kk} + 2M_r^\mu \epsilon_{ii} + \frac{1}{\omega_\ell}\sum_{\ell=1}^{L}[(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)\xi_\ell^{kk} + 2\mu Y_\ell^\mu \xi_\ell^{ii}].
\end{aligned} \tag{B.14}$$

Thus, we can write

$$\sigma_{ii} - (M_r^\kappa - \frac{2}{3}M_r^\mu)\epsilon_{kk} - \frac{1}{\omega_\ell}\sum_{\ell=1}^{L}[(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)\xi_\ell^{kk} + 2\mu Y_\ell^\mu \xi_\ell^{ii}] = 2M_r^\mu \epsilon_{ii} \tag{B.15}$$

$$\begin{aligned}
&\sum_i \langle 2M_r^\mu \dot{\epsilon}_{ii}, \epsilon_{ii}\rangle_\Omega = \sum_i \langle \dot{\epsilon}_{ii}, \sigma_{ii} - (M_r^\kappa - \frac{2}{3}M_r^\mu)\epsilon_{kk} - \frac{1}{\omega_\ell}\sum_{\ell=1}^{L}[(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)\xi_\ell^{kk} + 2\mu Y_\ell^\mu \xi_\ell^{ii}]\rangle_\Omega \\
&= \sum_i \langle \dot{\epsilon}_{ii}, \sigma_{ii}\rangle_\Omega - \langle (M_r^\kappa - \frac{2}{3}M_r^\mu)\epsilon_{kk}, \dot{\epsilon}_{kk}\rangle_\Omega - \sum_{\ell=1}^{L}\langle \frac{(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)}{\omega_\ell}\xi_\ell^{kk}, \dot{\epsilon}_{kk}\rangle_\Omega - \sum_{\ell=1}^{L}\sum_i \langle \frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ii}, \dot{\epsilon}_{ii}\rangle_\Omega
\end{aligned} \tag{B.16}$$

$$\langle (M_r^\kappa - \frac{2}{3}M_r^\mu)\epsilon_{kk}, \dot{\epsilon}_{kk}\rangle_\Omega + \sum_i \langle 2M_r^\mu \dot{\epsilon}_{ii}, \epsilon_{ii}\rangle_\Omega = \sum_i \langle \dot{\epsilon}_{ii}, \sigma_{ii}\rangle_\Omega - \sum_{\ell=1}^{L}\langle \frac{(\kappa Y_\ell^\kappa - \frac{2}{3}\mu Y_\ell^\mu)}{\omega_\ell}\xi_\ell^{kk}, \dot{\epsilon}_{kk}\rangle_\Omega - \sum_{\ell=1}^{L}\sum_i \langle \frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ii}, \dot{\epsilon}_{ii}\rangle_\Omega. \tag{B.17}$$

According to equations (A.13) and (B.8), setting $i \neq j$ gives

$$\begin{aligned}
\sigma_{ij} = 2\mu\epsilon_{ij} - 2\mu\sum_{\ell=1}^{L} Y_\ell^\mu \zeta_\ell^{ij} =& 2\mu\epsilon_{ij} - 2\mu\sum_{\ell=1}^{L} Y_\ell^\mu(-\frac{1}{\omega_\ell}\partial_t \zeta_\ell^{ij} + \epsilon_{ij}) = 2M_r^\mu \epsilon_{ij} - \sum_{\ell=1}^{L}\frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij}, \\
\Rightarrow \sigma_{ij} + \sum_{\ell=1}^{L}\frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij} =& 2M_r^\mu \epsilon_{ij}.
\end{aligned} \tag{B.18}$$

Taking the innner product over Ω with $\frac{1}{2M_r^\mu}(\sigma_{ij} - \sum_{\ell=1}^{L}\frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij})$ gives

$$\langle 2M_r^\mu \epsilon_{ij}, \dot{\epsilon}_{ij}\rangle_\Omega = \langle (\dot{\epsilon}_{ij}, \sigma_{ij}\rangle_\Omega - \sum_{\ell=1}^{L}\langle \dot{\epsilon}_{ij}, \frac{2\mu Y_\ell^\mu}{\omega_\ell}\xi_\ell^{ij}\rangle_\Omega \tag{B.19}$$

$$\Rightarrow \sum_{i\neq j}\langle 2M_r^{\mu}\epsilon_{ij},\dot{\epsilon}_{ij}\rangle_\Omega = \sum_{i\neq j}\langle(\dot{\epsilon}_{ij},\sigma_{ij}\rangle_\Omega - \sum_{\ell=1}^{L}\sum_{i\neq j}\langle\dot{\epsilon}_{ij},\frac{2\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{ij}\rangle_\Omega. \tag{B.20}$$

Summing over equations (B.17) and (B.20) gives

$$\begin{aligned}
&\langle(M_r^{\kappa}-\frac{2}{3}M_r^{\mu})\epsilon_{kk},\dot{\epsilon}_{kk}\rangle_\Omega + \sum_i\sum_j\langle 2M_r^{\mu}\epsilon_{ij},\dot{\epsilon}_{ij}\rangle_\Omega\\
=&\langle(M_r^{\kappa}-\frac{2}{3}M_r^{\mu})\epsilon_{kk},\dot{\epsilon}_{kk}\rangle_\Omega + \sum_i\langle 2M_r^{\mu}\dot{\epsilon}_{ii},\epsilon_{ii}\rangle_\Omega + \sum_{i\neq j}\langle 2M_r^{\mu}\epsilon_{ij},\dot{\epsilon}_{ij}\rangle_\Omega\\
=&\sum_i\langle\dot{\epsilon}_{ii},\sigma_{ii}\rangle_\Omega - \sum_{\ell=1}^{L}\langle\frac{(\kappa Y_\ell^{\kappa}-\frac{2}{3}\mu Y_\ell^{\mu})}{\omega_\ell}\xi_\ell^{kk},\dot{\epsilon}_{kk}\rangle_\Omega - \sum_{\ell=1}^{L}\sum_i\langle\frac{2\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{ii},\dot{\epsilon}_{ii}\rangle_\Omega\\
&+\sum_{i\neq j}\langle(\dot{\epsilon}_{ij},\sigma_{ij}\rangle_\Omega - \sum_{\ell=1}^{L}\sum_{i\neq j}\langle\dot{\epsilon}_{ij},\frac{2\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{ij}\rangle_\Omega\\
=&\sum_i\sum_j\langle\dot{\epsilon}_{ij},\sigma_{ij}\rangle_\Omega - \sum_{\ell=1}^{L}\langle\frac{(\kappa Y_\ell^{\kappa}-\frac{2}{3}\mu Y_\ell^{\mu})}{\omega_\ell}\xi_\ell^{kk},\dot{\epsilon}_{kk}\rangle_\Omega - \sum_{\ell=1}^{L}\sum_i\sum_j\langle\frac{2\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{ij},\dot{\epsilon}_{ij}\rangle_\Omega.
\end{aligned} \tag{B.21}$$

Combining equation (B.21) with the relationship of memory variables in (B.12) and (B.13), we end up with

$$\begin{aligned}
\partial_t E_s =&\langle(M_r^{\kappa}-\frac{2}{3}M_r^{\mu})\epsilon_{kk},\dot{\epsilon}_{kk}\rangle_\Omega + \sum_i\sum_j\langle 2M_r^{\mu}\epsilon_{ij},\dot{\epsilon}_{ij}\rangle_\Omega\\
&+\sum_{\ell=1}^{L}\sum_i\sum_j\langle\frac{2\mu Y_\ell^{\mu}}{\omega_\ell^2}\partial_t\xi_\ell^{ij},\xi_\ell^{ij}\rangle_\Omega + \sum_{\ell=1}^{L}\langle\frac{\kappa Y_\ell^{\kappa}-\frac{2}{3}\mu Y_\ell^{\mu}}{\omega_\ell^2}\partial_t\xi_\ell^{kk},\xi_\ell^{kk}\rangle_\Omega\\
&=\sum_i\sum_j\langle\dot{\epsilon}_{ij},\sigma_{ij}\rangle_\Omega - \sum_{\ell=1}^{L}\sum_i\sum_j\langle\frac{2\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{ij},\xi_\ell^{ij}\rangle_\Omega - \sum_{\ell=1}^{L}\langle\frac{\kappa Y_\ell^{\kappa}-\frac{2}{3}\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{kk},\xi_\ell^{kk}\rangle_\Omega.
\end{aligned} \tag{B.22}$$

We may expand the first term above to reveal its relation to the kinetic energy in (B.6) through

$$\begin{aligned}
&\sum_i\sum_j\langle\dot{\epsilon}_{ij},\sigma_{ij}\rangle_\Omega = \sum_i\langle\dot{\epsilon}_{ii},\sigma_{ii}\rangle_\Omega + \sum_{i\neq j}\langle\dot{\epsilon}_{ij},\sigma_{ij}\rangle_\Omega\\
=&\sum_i\langle v_{i,i},\sigma_{ii}\rangle_\Omega + \frac{1}{2}\sum_{i\neq j}\langle v_{i,j}+v_{j,i},\sigma_{ij}\rangle_\Omega = \sum_i\langle v_{i,i},\sigma_{ii}\rangle_\Omega + \frac{1}{2}\sum_{i\neq j}\langle v_{i,j},\sigma_{ij}\rangle_\Omega + \frac{1}{2}\sum_{i\neq j}\langle v_{j,i},\sigma_{ij}\rangle_\Omega\\
=&\sum_i\langle v_{i,i},\sigma_{ii}\rangle_\Omega + \frac{1}{2}\sum_{i\neq j}\langle v_{i,j},\sigma_{ij}\rangle_\Omega + \frac{1}{2}\sum_{i\neq j}\langle v_{j,i},\sigma_{ji}\rangle_\Omega(\sigma_{ij}=\sigma_{ji})\\
=&\sum_i\langle v_{i,i},\sigma_{ii}\rangle_\Omega + \sum_{i\neq j}\langle v_{i,j},\sigma_{ij}\rangle_\Omega = \sum_i\sum_j\langle v_{i,j},\sigma_{ij}\rangle_\Omega = -\sum_i\sum_j\langle v_i,\sigma_{ij,j}\rangle_\Omega\\
=&-\sum_i\langle\rho\partial_t v_i,v_i\rangle_\Omega = -\langle\rho\partial_t\mathbf{v},\mathbf{v}\rangle_\Omega = -\partial_t E_k.
\end{aligned} \tag{B.23}$$

As a result, we conclude that

$$\partial_t E = \partial_t E_k + \partial_t E_s = -\sum_{\ell=1}^{L}\sum_i\sum_j\langle\frac{2\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{ij},\xi_\ell^{ij}\rangle_\Omega - \sum_{\ell=1}^{L}\langle\frac{\kappa Y_\ell^{\kappa}-\frac{2}{3}\mu Y_\ell^{\mu}}{\omega_\ell}\xi_\ell^{kk},\xi_\ell^{kk}\rangle_\Omega \le 0. \tag{B.24}$$

Therefore, the viscoelastic sytem is also stable because the total energy is a monotonic decreasing function of time.

APPENDIX C: EQUIVALENCE AMONG DIFFERENT ADJOINT FORMULATIONS

It is interesting to note that the adjoint equation (64) looks slightly different from the forward equation (46). Omitting the sources, we may take another time derivative for the first-order forward system, which yields

$$\rho\partial_{tt}\mathbf{v} = DCD^T\mathbf{v} - D(C::\Gamma)\sum_{\ell=1}^{L}y_\ell\boldsymbol{\xi}_\ell = DCD^T\mathbf{v} - D\sum_{\ell=1}^{L}\partial_t\boldsymbol{\eta}_\ell \tag{C.1}$$

where we introduce the new variables $\partial_t\boldsymbol{\eta}_\ell = (C::\Gamma)y_\ell\boldsymbol{\xi}_\ell$, which are linear combinations of different components of the vector ξ_ℓ weighted by $C::\Gamma$. The 2nd order time derivative for η_ℓ is

$$\begin{aligned}
\partial_{tt}\boldsymbol{\eta}_\ell =&(C::\Gamma)y_\ell\partial_t\boldsymbol{\xi}_\ell = (C::\Gamma)y_\ell\omega_\ell(-\boldsymbol{\xi}_\ell + D^T\mathbf{v})\\
=&-\omega_\ell\partial_t\boldsymbol{\eta}_\ell + y_\ell\omega_\ell(C::\Gamma)D^T\mathbf{v}.
\end{aligned} \tag{C.2}$$

which is equivalent to

$$\begin{cases}\rho\partial_{tt}\mathbf{v} = DCD^T\mathbf{v} - D\sum_{\ell=1}^{L}\partial_t\boldsymbol{\eta}_\ell \\ \partial_{tt}\boldsymbol{\eta}_\ell = -\omega_\ell\partial_t\boldsymbol{\eta}_\ell + y_\ell\omega_\ell(C :: \Gamma)D^T\mathbf{v}.\end{cases} \tag{C.3}$$

Similarly, we may take the second-order time derivative and eliminate the stress in the adjoint system, yielding

$$\begin{cases}\rho\partial_{tt}\bar{\mathbf{v}} = DCD^T\bar{\mathbf{v}} + D\sum_{\ell=1}^{L}\partial_t\bar{\boldsymbol{\xi}}_\ell \\ \partial_{tt}\bar{\boldsymbol{\xi}}_\ell = \omega_\ell\partial_t\bar{\boldsymbol{\xi}}_\ell + y_\ell\omega_\ell(C :: \Gamma)D^T\bar{\mathbf{v}}.\end{cases} \tag{C.4}$$

Keeping in mind that the adjoint equation has to be integrated from the final time ($t = T$) to the starting time ($t = 0$), we define the quantities $\bar{\boldsymbol{\xi}}'_\ell(T-t) = \bar{\boldsymbol{\xi}}_\ell(t)$, $\bar{\mathbf{v}}'(T-t) = \bar{\mathbf{v}}(t)$ such that

$$\partial_t\bar{\boldsymbol{\xi}}_\ell = -\partial_t\bar{\boldsymbol{\xi}}'_\ell, \partial_{tt}\bar{\boldsymbol{\xi}}_\ell = \partial_{tt}\bar{\boldsymbol{\xi}}'_\ell, \partial_{tt}\bar{\mathbf{v}}' = \partial_{tt}\bar{\mathbf{v}}. \tag{C.5}$$

Then, the adjoint system can be expressed as

$$\begin{cases}\rho\partial_{tt}\bar{\mathbf{v}}' = DCD^T\bar{\mathbf{v}}' - D\sum_{\ell=1}^{L}\partial_t\bar{\boldsymbol{\xi}}_\ell \\ \partial_{tt}\bar{\boldsymbol{\xi}}'_\ell = -\omega_\ell\partial_t\bar{\boldsymbol{\xi}}'_\ell + y_\ell\omega_\ell(C :: \Gamma)D^T\bar{\mathbf{v}}',\end{cases} \tag{C.6}$$

which implies that the adjoint wave equation ends up with the same solution as the forward wave equation according to its 2nd order expression; the Green function are the same except the time directions are opposite. This verifies the equivalence of the different adjoint systems, including the one presented in this paper, the one from Fichtner & van Driel (2014), and the Green function manipulation in Tarantola (1988); Charara et al. (2000), as well as the second order formulation by Tromp et al. (2005).