Internship report for the Instituto Colombia del Petróleo (ICP)
In requirement of the degree Master of science Applied Physics, TU Delft, the Netherlands

Full waveform inversion of isotropic elastic media using an adjoint state method in finite difference time-domain

J. J. Wesdorp*

Abstract
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Supervisors
Herling Gonzalez-Alvarez [1]
Prof. Koen van Dongen [2]
Prof. Ana Ramirez [3]

[1] *Instituto Colombia del Petróleo (ICP), Piedecuesta, Colombia*
[2] *Technical University of Delft, the Netherlands*
[3] *Universidad Industrial de Santander (UIS), CPS research group, Bucaramanga, Colombia*
***email**: jaapwesdorp@gmail.com

Contents

1. Introduction

Full waveform inversion can be used a tool to quantitatively image material properties of the earth, or tissue.

30 years ago -> [1] proposed FWI. Now since 10 years the industry is making use of FWI in their processing workflow due to the gain in computing power.

Initially mostly acoustic, but there is interest in modeling elastic waves, since you can see phenomena like: groundroll waves, surface waves ->(WHAT ELSE IS SPECIAL ABOUT ELASTIC).

FWI requires one to know the gradient of your guessed model with respect to measured shots, therefore the *adjoint state method* [2] makes this a lot more efficient.

This research was performed as part of an internship for the Instituto Colombia del Petróleo (ICP). The first goal was to model the isotropic 2D elastic wave equation, which is described in section ... The second goal was to find the adjoint operator and gradients for the elastic case. Since no explicit declaration is given yet for the elastic case in current literature and the adjoint state method is usually described in a very abstract way, this work shows an "engineers approach" to obtaining the adjoint system and gradient expressions for any general set of equations in sec .... This method is applied to the elastic case in sec ..., and more examples are given in Appendix B. This method is subsequently tested by performing FWI on a test model described in sec ...

2. A simple recipe for the adjoint state method

2.1 Adjoint method for general discretized linear systems

This section derives in simple steps how to obtain the adjoint equations and gradients given a certain system of equations. These are cast in a general form usually present in acoustic, elastic or electromagnetic FWI applications. The proof of validity of this method is presented in Appendix A, in which we derived the method for an more general set of equations based on the notes of Bradley[3].

2.1.1 Formulation

Say we have a model representing the earth (in seismic applications) or tissue (in medical applications) for some parameter like density ρ or speed of sound v, which we discretized in a grid in space into $N_g (= N_x \times N_z$ in 2D) points. If we want to model d_m parameters over this grid we can write the values of these parameters over the grid as a vector $\vec{m}$ of dimension $d_m \times N_g$. We want to model d_s fields also discretized in space on the same grid we can write them as a vector $\vec{s}$ of dimension $d_s \times N_g$. useful simplification for already (only in space!) discretized wave equations(e.g electromagnetic, elastic or acoustic), whether written as a set of multiple first or second order PDE's in time and space, is the following form

$$
\begin{aligned}
\min_{\vec{m}} \; & \chi(\vec{s},\vec{m}) = \int_0^T f(\vec{s},\vec{m})\, dt \\
\text{subject to} \quad & T(\vec{m})\ddot{\vec{s}} - C(\vec{m})\dot{\vec{s}} - A(\vec{m})\vec{s} - b(\vec{m}) = 0 \\
\text{with B.C} \quad & \vec{s}(0) = 0 \\
& \dot{\vec{s}}(0) = 0
\end{aligned}
\tag{1}
$$

where the T, C, A are big matrices of coefficients which follow from the discretization in space of a PDE system and $b(\vec{m})$ contains the source terms.

2.1.2 Adjoint formulation

Using Eq. (28) and applying it to our system of Eq. (1) we obtain

$$
T^T(\vec{m})\ddot{\vec{\lambda}} = -C^T(\vec{m})\dot{\vec{\lambda}} + A^T(\vec{m})\vec{\lambda} - \frac{\partial f}{\partial \vec{s}}
\tag{2}
$$

2.1.3 Gradients

We can apply Eq. (29) to the system of Eq. (1) and find the following expression for the total gradient with respect to *all* model parameters

$$\frac{d\chi}{d\vec{m}} = \int_0^T \frac{\partial f}{\partial \vec{m}} + \vec{\lambda}^T \left(\frac{\partial T}{\partial \vec{m}} \ddot{\vec{s}} - \frac{\partial C}{\partial \vec{m}} \dot{\vec{s}} - \frac{\partial A}{\partial \vec{m}} \vec{s} - \frac{\partial b}{\partial \vec{m}} \right) dt \tag{3}$$

Note that the gradients $\frac{\partial T}{\partial \vec{m}}, \frac{\partial C}{\partial \vec{m}}, \frac{\partial A}{\partial \vec{m}}$ can be seen as a sum of all the $3N$ derivatives of each component of the matrices $T(\vec{m}), C(\vec{m}), A(\vec{m})$ to the discrete variables $\vec{m}_i$, but in practice we can write this into simple compact expressions as shown in the examples below.

2.1.4 Least squares cost function

In most cases the cost function is given by the misfit between observed data d at certain receiver positions y_R and the modeled data at the same positions s_{y_R}, so $f(\vec{s})$ is given by

$$f(\vec{s}) = \frac{1}{2} \sum_R^{N_R} ||s_{y_R} - d_{y_R}||^2 \tag{4}$$

where the sum goes over all receiver positions. Note that $f(\vec{s})$ does not depend directly on the model parameters $\vec{m}$ making the gradient expression simpler since $\frac{\partial f}{\partial \vec{m}} = 0$. We thus obtain

$$\frac{\partial f}{\partial \vec{s}} = \sum_R^{N_R} ||s_{y_R} - d_{y_R}|| \tag{5}$$

2.2 Adjoint and gradients of the elastic wave equation

Elastic waves in 2D isotropic media can be modeled by using only three parameters, the density ρ and the two Lame parameters μ, λ and follow the following system of equations

$$\begin{aligned}
\rho \frac{\partial v_x}{\partial t} &= \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xz}}{\partial z} \\
\rho \frac{\partial v_z}{\partial t} &= \frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{zz}}{\partial z} \\
\frac{\partial \tau_{xx}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_x}{\partial x} + \lambda \frac{\partial v_z}{\partial z} + b_{xx} \\
\frac{\partial \tau_{zz}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_z}{\partial z} + \lambda \frac{\partial v_x}{\partial x} + b_{zz} \\
\frac{\partial \tau_{xz}}{\partial t} &= \mu \left(\frac{\partial v_x}{\partial z} \frac{\partial v_z}{\partial x} \right)
\end{aligned} \tag{6}$$

where we only added the source to the stress as to resemble an earthquake. The form of b is independent of the model parameters and usually taken as a ricker wavelet. writing this in the form of Eq. (1) gives $T(m) = 0$ and

$$s = \begin{bmatrix} v_x \\ v_z \\ \tau_{xx} \\ \tau_{xz} \\ \tau_{zx} \end{bmatrix},\ C(m) = -\begin{bmatrix} \vec{\rho} & 0 & 0 & 0 & 0 \\ 0 & \vec{\rho} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

and

$$A(m) = \begin{bmatrix} 0 & 0 & \mathscr{D}_x & 0 & \mathscr{D}_z \\ 0 & 0 & 0 & \mathscr{D}_z & \mathscr{D}_x \\ (\lambda+2\mu)\mathscr{D}_x & \lambda\mathscr{D}_z & 0 & 0 & 0 \\ \lambda\mathscr{D}_x & (\lambda+2\mu)\mathscr{D}_z & 0 & 0 & 0 \\ \mu\mathscr{D}_z & \mu\mathscr{D}_x & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

Thus we have $C^T(m) = C(m)$ and

$$A^T(m) = \begin{bmatrix} 0 & 0 & -\mathscr{D}_x(\lambda+2\mu) & -\mathscr{D}_x\lambda & -\mathscr{D}_z\mu \\ 0 & 0 & -\mathscr{D}_z\lambda & -\mathscr{D}_z(\lambda+2\mu) & -\mathscr{D}_x\mu \\ -\mathscr{D}_x & 0 & 0 & 0 & 0 \\ 0 & -\mathscr{D}_z & 0 & 0 & 0 \\ -\mathscr{D}_z & -\mathscr{D}_x & 0 & 0 & 0 \end{bmatrix} \tag{9}$$

which in turn (when transforming back to the continuous domain from the discrete variables) results in the following adjoint equations(naming the adjoint variable l) as defined in Eq. (2)

$$\begin{aligned} \rho\frac{\partial l_1}{\partial t} &= \frac{\partial((\lambda+2\mu)l_3)}{\partial x} + \frac{\partial(\lambda l_4)}{\partial x} + \frac{\partial(\mu\lambda_5)}{\partial z} \\ \rho\frac{\partial l_2}{\partial t} &= \frac{\partial(\lambda l_3)}{\partial z} + \frac{\partial((\lambda+2\mu)l_4)}{\partial z} + \frac{\partial(\mu\lambda_5)}{\partial x} \\ \frac{\partial l_3}{\partial t} &= \frac{\partial l_1}{\partial x} \\ \frac{\partial l_4}{\partial t} &= \frac{\partial l_2}{\partial z} \\ \frac{\partial l_5}{\partial t} &= \frac{\partial l_1}{\partial z} + \frac{\partial l_2}{\partial x} \end{aligned} \tag{10}$$

where the order of differentiation and multiplication with the model parameters is important. The gradients then follow from Eq. (3), giving

$$
\begin{aligned}
\frac{d\chi}{d\rho} &= \int_0^T -l^T \frac{\partial C}{\partial \rho} \dot{s}\, dt \\
&= \int_0^T \begin{bmatrix} \vec{l}_1 & \vec{l}_2 & \vec{l}_3 & \vec{l}_4 & \vec{l}_5 \end{bmatrix} \begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\vec{v}}_x \\ \dot{\vec{v}}_y \\ \dot{\vec{\tau}}_{xx} \\ \dot{\vec{\tau}}_{zz} \\ \dot{\vec{\tau}}_{xz} \end{bmatrix} \\
&= \int_0^T \vec{l}_1 \dot{\vec{v}}_x + \vec{l}_2 \dot{\vec{v}}_z \, dt
\end{aligned} \tag{11}
$$

,

$$
\begin{aligned}
\frac{d\chi}{d\lambda} &= \int_0^T -l^T \frac{\partial A}{\partial \lambda} s\, dt \\
&= -\int_0^T \begin{bmatrix} \vec{l}_1 & \vec{l}_2 & \vec{l}_3 & \vec{l}_4 & \vec{l}_5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \mathscr{D}_x & \mathscr{D}_z & 0 & 0 & 0 \\ \mathscr{D}_x & \mathscr{D}_z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{v}_x \\ \vec{v}_y \\ \vec{\tau}_{xx} \\ \vec{\tau}_{zz} \\ \vec{\tau}_{xz} \end{bmatrix} \\
&= -\int_0^T \left(\vec{l}_3 + \vec{l}_4 \right) \left(\frac{\partial \vec{v}_x}{\partial x} + \frac{\partial \vec{v}_z}{\partial z} \right)
\end{aligned} \tag{12}
$$

and

$$
\begin{aligned}
\frac{d\chi}{d\mu} &= \int_0^T -l^T \frac{\partial A}{\partial \mu} s\, dt \\
&= -\int_0^T \begin{bmatrix} \vec{l}_1 & \vec{l}_2 & \vec{l}_3 & \vec{l}_4 & \vec{l}_5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2\mathscr{D}_x & 0 & 0 & 0 & 0 \\ 0 & 2\mathscr{D}_z & 0 & 0 & 0 \\ \mathscr{D}_z & \mathscr{D}_x & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{v}_x \\ \vec{v}_y \\ \vec{\tau}_{xx} \\ \vec{\tau}_{zz} \\ \vec{\tau}_{xz} \end{bmatrix} \\
&= -\int_0^T 2\vec{l}_3 \frac{\partial \vec{v}_x}{\partial x} + 2\vec{l}_4 \frac{\partial \vec{v}_z}{\partial z} + \vec{l}_5 \left(\frac{\partial \vec{v}_x}{\partial z} + \frac{\partial \vec{v}_z}{\partial x} \right) dt
\end{aligned} \tag{13}
$$

3. Numerical implementation

Chapter intro here

3.1 Staggered grid discrete equations

We now show the modeling of Eq. (6). Due to the nature of the equations, if we would implement every field at the same grid point in time and space using centered finite differences, we would get decoupling between the velocity and stress in time and space(e.g V_x at time t depending on τ at time $t \pm \frac{dt}{2}$ which we do not have defined). This could be solved by using a two times finer grid, but can be more elegantly solved by use a staggered grid following [4]. Using the notation $\{v_x, v_z, \tau_{xx}, \tau_{zz}, \tau_{xz}\} = \{U, V, X, Z, T\}$ we do a "leapfrog" scheme, which in each timestep loops over the spatial directions for calculating $U^{n+\frac{1}{2}}, V^{n+\frac{1}{2}}$ as a function of the previously calculated X^n, Z^n, T^n first and consecutively loops over the spatial directions again to calculate $X^{n+1}, Z^{n+1}, T^{n+1}$ depending on the just calculated $U^{n+\frac{1}{2}}, V^{n+\frac{1}{2}}$. This avoids the decoupling mentioned above, but requires the fields to be defined in different points in the grid. The numerical equations are then given by (using a 2nd order central difference in time and arbitrary difference operator in space)

$$
\begin{cases}
U_{i,j}^{n+\frac{1}{2}} = U_{i,j}^{n-\frac{1}{2}} + \dfrac{\Delta t}{\rho_{i,j}} \left[D_x X_{i,j}^n + D_z T_{i,j}^n \right] \\
V_{i+\frac{1}{2},j+\frac{1}{2}}^{n+\frac{1}{2}} = V_{i+\frac{1}{2},j+\frac{1}{2}}^{n-\frac{1}{2}} + \dfrac{\Delta t}{\rho_{i+\frac{1}{2},j+\frac{1}{2}}} \left[D_x T_{i+\frac{1}{2},j+\frac{1}{2}}^n + D_z Z_{i+\frac{1}{2},j+\frac{1}{2}}^n \right] \\
X_{i+\frac{1}{2},j}^{n+1} = X_{i+\frac{1}{2},j}^n + \Delta t \left[\left(\lambda_{i+\frac{1}{2},j} + 2\mu_{i+\frac{1}{2},j} \right) D_x U_{i+\frac{1}{2},j}^{n+\frac{1}{2}} + \lambda_{i+\frac{1}{2},j} D_z V_{i+\frac{1}{2},j}^{n+\frac{1}{2}} \right] \\
Z_{i+\frac{1}{2},j}^{n+1} = Z_{i+\frac{1}{2},j}^n + \Delta t \left[\left(\lambda_{i+\frac{1}{2},j} + 2\mu_{i+\frac{1}{2},j} \right) D_z V_{i+\frac{1}{2},j}^{n+\frac{1}{2}} + \lambda_{i+\frac{1}{2},j} D_x U_{i+\frac{1}{2},j}^{n+\frac{1}{2}} \right] \\
T_{i,j+\frac{1}{2}}^{n+1} = T_{i,j+\frac{1}{2}}^n + \mu_{i,j+\frac{1}{2}} \Delta t \left[D_z U_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} + D_x V_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} \right]
\end{cases}
\tag{14}
$$

where for the simplest second order differential operator $D_x A_{i,j}^n = \frac{A_{i+\frac{1}{2},j}^n - A_{i-\frac{1}{2},j}^n}{\Delta_x}$ and $D_z A_{i,j}^n = \frac{A_{i,j+\frac{1}{2}}^n - A_{i,j-\frac{1}{2}}^n}{\Delta_z}$. n denotes the discrete time step, i, j the x, z coordinates respectively, and $\Delta t, \Delta x, \Delta z$ are the stepsizes taken into each direction respectively.

Note that in the actual implementation all arrays start at index 0. Therefore the indexes of U remain just i, j, but for the others we get

$$
\begin{cases}
U: & i \rightarrow i',\ j \rightarrow j' \\
V: & i + \dfrac{1}{2} \rightarrow i',\ j + \dfrac{1}{2} \rightarrow j' \\
X: & i + \dfrac{1}{2} \rightarrow i',\ j \rightarrow j' \\
Z: & i + \dfrac{1}{2} \rightarrow i',\ j \rightarrow j' \\
T: & i \rightarrow i',\ j + \dfrac{1}{2} \rightarrow j'
\end{cases}
\tag{15}
$$

where $'$ stands for the index in the code, since we imagine them displaced from the origin.

3.2 CPML Absorbing boundaries

In most seismic cases we have an infinite domain of our model, but we are just interested in a specific part of this domain. In order to simulate only this part without reflecting boundaries one needs to simulate "open" boundaries. This can be done by introducing an absorbing layer at the boundaries originally introduced as a Perfectly Matched Layer(PML) [6] but later improved to a Convolutionary Perfectly Matched Layer(CPML) [5] in order to reduce reflection for waves at grazing angle with the layer.

CPML consists changing the spatial derivatives seen in the wave equations inside the layer by adding an imaginary part.

In this section we introduce the widely used convolutional perfectly matched layer(CPML) boundary conditions [5], which are an improvement of the PML proposed 20 years earlier [6]. This simulates having an infinite domain(which is numerically infeasible) by absorbing waves that reach the boundaries.

3.3 Obtaining realistic model parameters

Modeling elastic waves requires knowledge of all three model parameters: the density ρ, and lamé parameters μ and λ. Usually in literature, instead of the Lamé parameters, the S and P-wave velocities are given since these can be read more inituitively. There are simple relations between the wavespeeds and Lamé parameters following from the elastic wave equations

$$\{\rho,\mu,\lambda\} \rightarrow \{\rho, v_s, v_p\}, \text{ with } v_s = \sqrt{\frac{\mu}{\rho}},\ v_p = \sqrt{\frac{\lambda+2\mu}{\rho}} \tag{16}$$

3.3.1 Brocher relations

Usually seismic tomography and refraction studies of the earth crust report only the P-wave velocity v_p, while we need the density and S-wave velocity as well. In order to obtain realistic estimates of these parameters we can the relations derived by T. M. Brocher in 2005 [7] relating v_p to v_s and v_p to ρ. These are derived from thousands of real world datasets of the earth's crust giving an empirical method to obtain ρ and v_s given known v_p(in km/s)

$$\rho[g/cm^3] = 1.6612v_p - 0.4721v_p^2 + 0.0671v_p^3 - 0.0043v_p^4 + 0.000106v_p^5 \tag{17}$$

and

$$v_s[km/s] = 0.7857 - 1.2344v_p + 0.7949v_p^2 - 0.1238v_p^3 + 0.0064v_p^4 \tag{18}$$

which are only valid between 1.5 km/s $< v_p < 8$ km/s

3.4 Total system energy

In any wave equation modeling, an important check is to measure the system energy, and see that if there is no source, energy remains constant in the system. For elastic waves the energy is given by a combination of kinetic energy K of the moving particles and spring energy T stored by displacing these particles from their equilibrium position following from Hooke's law.

$$E = K + T = \frac{1}{2}\rho ||v||^2 + \frac{1}{2}\sum \sigma_{ij}\varepsilon_{ij} \tag{19}$$

4. Elastic simulation results

5. FWI results

6. Conclusion

Appendices

1. Lagrangian based derivation of the gradient via the adjoint state method

In this section the adjoint equations and the gradients for a general first or second order PDE system will be derived following the approach outlined in [3].

A.1 General system of equations

We can write the most general minimization problem of a cost function subject any first order or second order (in time) set of partial differential equations(PDE) in the following form

$$\begin{aligned} \min_{m} \quad & \chi(s,m) = \int_0^T f(s,m)\,dt \\ \text{subject to} \quad & h(\ddot{s},\dot{s},s,m,t) = 0 \\ \text{with B.C} \quad & g((s(0),m) = 0 \\ & k((\dot{s}(0),m) = 0 \end{aligned} \tag{20}$$

where s is a discretized vector of the fields and the dots denote derivatives to time, m is the vector of all the discretized model parameters, t denotes time, $h(\ddot{s},\dot{s},s,m,t) = 0$ is the system of equations, e.g the wave equation in 2D or 3D,T is the final time of integration and $g((s(0),m), k((\dot{s}(0),m)$ denote initial conditions for the field vector s and its derivative $\dot{s}$. Note that this is similar to [3] but with the added explicit dependence of both $\dot{s}$ and $\ddot{s}$ giving a more general expression for the adjoint equations and the gradients.

A.2 Derivative to model parameters

When solving the minimization problem numerically, one often uses a method similar to Newton's gradient descend. This thus requires knowledge of the derivative of the cost function $\chi(s,m)$ to all of the model parameters m. Using the chain rule we obtain:

$$\frac{d\chi}{dm} = \frac{\partial\chi}{\partial s}\frac{\partial s}{\partial m} + \frac{\partial\chi}{\partial m} \tag{21}$$

This depends on the Frechet derivatives $\frac{\partial s}{\partial m}$ which require at least $3N$ evaluations of the forward model in order to obtain an estimate, we thus want to avoid the calculation of this term. To facilitate this we define the Lagrangian $\mathscr{L}$ by

$$\mathscr{L} = \int_0^T \left[f(s,m) + \lambda^T h(\ddot{s},\dot{s},s,m,t)\right] dt + \mu^T g((s(0),m) + \eta^T k((\dot{s}(0),m) \tag{22}$$

where the auxiliary variables λ,μ,η have the same length as the discretized field vector s. Note that due to the initial conditions of Eq. (20) we have $\frac{d\chi}{dm} = \frac{d\mathscr{L}}{dm}$, where using the chain rule repetitively

$$\begin{aligned}\frac{d\mathscr{L}}{dm} = \int_0^T &\left[\frac{\partial f}{\partial s}\frac{\partial s}{\partial m} + \frac{\partial f}{\partial m}\right.\\ &\left.+\lambda^T\left(\frac{\partial h}{\partial \ddot{s}}\frac{\partial \ddot{s}}{\partial m} + \frac{\partial h}{\partial \dot{s}}\frac{\partial \dot{s}}{\partial m} + \frac{\partial h}{\partial s}\frac{\partial s}{\partial m} + \frac{\partial h}{\partial m}\right)\right] dt\\ &+\mu^T\left(\frac{\partial g}{\partial s(0)}\frac{\partial s(0)}{\partial m} + \frac{\partial g}{\partial m}\right)\\ &+\eta^T\left(\frac{\partial k}{\partial \dot{s}(0)}\frac{\partial \dot{s}(0)}{\partial m} + \frac{\partial k}{\partial m}\right)\end{aligned} \tag{23}$$

This looks intimidating and not really much simpler, but we are still free to choose the expressions for λ,μ and η. We will choose this such that we can avoid calculating the computationally difficult derivatives $\frac{\partial s}{\partial m}$. But first we need to do some partial integration to get rid of $\frac{\partial \dot{s}}{\partial m}$ and $\frac{\partial \ddot{s}}{\partial m}$. With a single partial integration we can write

$$\int_0^T \lambda^T\frac{\partial h}{\partial \dot{s}}\frac{\partial \dot{s}}{\partial m} = \left[\lambda^T\frac{\partial h}{\partial \dot{s}}\frac{\partial s}{\partial m}\right]_0^T - \int_0^T \frac{\partial s}{\partial m}\left(\dot{\lambda}^T\frac{\partial h}{\partial \dot{s}} + \lambda^T\frac{\partial}{\partial t}\frac{\partial h}{\partial \dot{s}}\right) \tag{24}$$

and with a double partial integration we can write

$$\begin{aligned}\int_0^T \lambda^T\frac{\partial h}{\partial \ddot{s}}\frac{\partial \ddot{s}}{\partial m} = &\left[\lambda^T\frac{\partial h}{\partial \ddot{s}}\frac{\partial \dot{s}}{\partial m} - \frac{\partial s}{\partial m}\left(\dot{\lambda}^T\frac{\partial h}{\partial \ddot{s}} + \lambda^T\frac{\partial}{\partial t}\frac{\partial h}{\partial \ddot{s}}\right)\right]_0^T\\ &+\int_0^T \frac{\partial s}{\partial m}\left(\ddot{\lambda}^T\frac{\partial h}{\partial \ddot{s}} + 2\dot{\lambda}^T\frac{\partial}{\partial t}\frac{\partial h}{\partial \ddot{s}} + \lambda^T\frac{\partial^2}{\partial t^2}\frac{\partial h}{\partial \ddot{s}}\right) dt\end{aligned} \tag{25}$$

So filling this back into Eq. (23) and regrouping terms gives

$$
\begin{aligned}
\frac{d\mathscr{L}}{dm} = \int_0^T & \left[\frac{\partial s}{\partial m} \left(\frac{\partial f}{\partial s} + \ddot{\lambda}^T \frac{\partial h}{\partial \ddot{s}} + \dot{\lambda}^T \left(2 \frac{\partial}{\partial t} \frac{\partial h}{\partial \ddot{s}} - \frac{\partial h}{\partial \dot{s}} \right) \right.\right. \\
& \left.\left. + \lambda^T \left(\frac{\partial h}{\partial s} + \frac{\partial^2}{\partial t^2} \frac{\partial h}{\partial \ddot{s}} - \frac{\partial}{\partial t} \frac{\partial h}{\partial \dot{s}} \right) \right) + \frac{\partial f}{\partial m} + \lambda^T \frac{\partial h}{\partial m} \right] dt \\
& + \left(\mu^T \frac{\partial g}{\partial s(0)} - \lambda^T \frac{\partial h}{\partial \dot{s}} + \dot{\lambda}^T \frac{\partial h}{\partial \ddot{s}} + \lambda^T \frac{\partial}{\partial t} \frac{\partial h}{\partial \ddot{s}} \right) \frac{\partial s}{\partial m} \bigg|_0 \\
& + \left(\eta^T \frac{\partial h}{\partial \dot{s}(0)} - \lambda^T \frac{\partial h}{\partial \ddot{s}} \right) \frac{\partial \dot{s}}{\partial m} \bigg|_0 \\
& + \lambda^T \frac{\partial h}{\partial \ddot{s}} \frac{\partial \dot{s}}{\partial m} \bigg|_T + \left(\lambda^T \frac{\partial h}{\partial \dot{s}} - \dot{\lambda}^T \frac{\partial h}{\partial \ddot{s}} - \lambda^T \frac{\partial}{\partial t} \frac{\partial h}{\partial \ddot{s}} \right) \frac{\partial s}{\partial m} \bigg|_T \\
& + \mu^T \frac{\partial g}{\partial m} + \eta^T \frac{\partial k}{\partial m}
\end{aligned} \tag{26}
$$

we can then choose values for the auxiliary variables such that undesired terms drop out of $\frac{d\mathscr{L}}{dm}$. If we set

$$
\begin{aligned}
\mu^T &= \left(\dot{\lambda}(0) \frac{\partial h}{\partial \ddot{s}} + \lambda^T(0) \left(\frac{\partial}{\partial t} \frac{\partial h}{\partial \ddot{s}} - \frac{\partial h}{\partial \dot{s}} \right) \right) \left(\frac{\partial g}{\partial s(0)} \right)^{-1} \\
\eta^T &= \left(\lambda(0) \frac{\partial h}{\partial \ddot{s}} \right) \left(\frac{\partial k}{\partial \dot{s}(0)} \right)^{-1} \\
\lambda(T) &= 0 \\
\dot{\lambda}(T) &= 0
\end{aligned} \tag{27}
$$

and let λ satisfy the following so called *adjoint* equation

$$
\ddot{\lambda}^T \frac{\partial h}{\partial \ddot{s}} + \dot{\lambda}^T \left(2 \frac{\partial}{\partial t} \frac{\partial h}{\partial \ddot{s}} - \frac{\partial h}{\partial \dot{s}} \right) + \lambda^T \left(\frac{\partial h}{\partial s} + \frac{\partial^2}{\partial t^2} \frac{\partial h}{\partial \ddot{s}} - \frac{\partial}{\partial t} \frac{\partial h}{\partial \dot{s}} \right) = - \frac{\partial f}{\partial s} \tag{28}
$$

we can see that the $\frac{\partial s}{\partial m}$ terms drop out of $\frac{d\mathscr{L}}{m}$ and we remain for the *gradients* only with

$$
\frac{d\mathscr{L}}{dm} = \int_0^T \frac{\partial f}{\partial m} + \lambda^T \frac{\partial h}{\partial m} dt + \mu^T \frac{\partial g}{\partial m} + \eta^T \frac{\partial k}{\partial m} \tag{29}
$$

where the two rightmost terms are zero if the initial conditions do not depend directly on the model parameters m. Eq. (28) is called the *adjoint* equation to the original system of equations $h(\ddot{s}, \dot{s}, s, m, t)$

2. More example applications

B.1 The second order electromagnetic wave equation

Now for an example with second order terms as well we take the following second order electromagnetic wave equation

$$\varepsilon \frac{\partial^2 E_y}{\partial t^2} = \frac{\partial}{\partial x}\left(\frac{1}{\mu}\frac{\partial E_y}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\mu}\frac{\partial E_y}{\partial z}\right) - \sigma \frac{\partial E_y}{\partial t} \tag{30}$$

which models the same as **??**. We can see that now $s = \vec{E}_y$ a single field variable (which is still a vector of length N)

$$T(m) = \varepsilon,\ C(m) = -\sigma,\ A(m) = \mathscr{D}_x \frac{1}{\mu} \mathscr{D}_x + \mathscr{D}_z \frac{1}{\mu} \mathscr{D}_z \tag{31}$$

All of these operators are self-adjoint, thus $A^T(m) = A(m),\ C^T(m) = C(m),\ T^T(m) = T(m)$, we thus get for the adjoint equation

$$\varepsilon \frac{\partial^2 \lambda}{\partial t^2} = \frac{\partial}{\partial x}\left(\frac{1}{\mu}\frac{\partial \lambda}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\mu}\frac{\partial \lambda}{\partial z}\right) + \sigma \frac{\partial \lambda}{\partial t} \tag{32}$$

where there is only a sign change in front of σ due to Eq. (2). The gradients then become according to Eq. (3)

$$\begin{aligned}
\frac{d\chi}{d\varepsilon} &= \int_0^T \vec{\lambda}^T \ddot{\vec{E}}_y \, dt \\
\frac{d\chi}{d\mu} &= \int_0^T \vec{\lambda}^T \left(\frac{\partial}{\partial x}\left(\frac{1}{\mu^2}\frac{\partial \vec{E}_y}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\mu^2}\frac{\partial \vec{E}_y}{\partial z}\right) \right) dt \\
\frac{d\chi}{d\sigma} &= \int_0^T \vec{\lambda}^T \dot{\vec{E}}_y \, dt
\end{aligned} \tag{33}$$

B.2 The second order acoustic wave equation

The acoustic wave equation with variable density ρ and velocity c is given by

$$\frac{1}{\rho c^2}\frac{\partial^2 p}{\partial t^2} = \left(\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial p}{\partial z}\right) \right) \tag{34}$$

we thus get

$$T(m) = \frac{1}{\rho c^2},\ C(m) = 0,\ A(m) = \mathscr{D}_x \frac{1}{\rho} \mathscr{D}_x + \mathscr{D}_z \frac{1}{\rho} \mathscr{D}_z \tag{35}$$

which looks very similar to the electromagnetic case, again $A^T(m) = A(m),\ C^T(m) = C(m),\ T^T(m) = T(m)$, so following Eq. (2) the adjoint equation is the same as the original equation

$$\frac{1}{\rho c^2}\frac{\partial^2 \lambda}{\partial t^2} = \left(\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \lambda}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial \lambda}{\partial z}\right) \right) \tag{36}$$

According to Eq. (3) we then obtain

$$
\begin{aligned}
\frac{d\chi}{dc} &= -\frac{2}{\rho c^3}\int_0^T \vec{\lambda}^T \vec{p}\, dt \\
\frac{d\chi}{d\rho} &= \int_0^T \vec{\lambda}^T \left(-\frac{1}{\rho^2 c^2}\vec{p} + \frac{\partial}{\partial x}\left(\frac{1}{\rho^2}\frac{\partial \vec{p}}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho^2}\frac{\partial \vec{p}}{\partial z}\right)\right)\, dt
\end{aligned} \tag{37}
$$

B.3 The first order acoustic wave equation

The first order acoustic wave equation with variable density ρ and velocity c is given by

$$
\begin{aligned}
\frac{1}{\rho c^2}\frac{\partial p}{\partial t} &= -\frac{\partial v_x}{\partial x} - \frac{\partial v_z}{\partial z} \\
\frac{\partial v_x}{\partial t} &= -\frac{1}{\rho}\frac{\partial p}{\partial x} \\
\frac{\partial v_z}{\partial t} &= -\frac{1}{\rho}\frac{\partial p}{\partial z}
\end{aligned} \tag{38}
$$

this gives

$$
T(m) = 0,\; C(m) = -\begin{bmatrix} \frac{1}{\rho c^2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\; A(m) = \begin{bmatrix} 0 & -\mathscr{D}_x & -\mathscr{D}_z \\ -\frac{1}{\rho}\mathscr{D}_x & 0 & 0 \\ -\frac{1}{\rho}\mathscr{D}_z & 0 & 0 \end{bmatrix} \tag{39}
$$

we thus get $C^T(m) = C$ and

$$
A^T(m) = \begin{bmatrix} 0 & \mathscr{D}_x\frac{1}{\rho} & \mathscr{D}_z\frac{1}{\rho} \\ \mathscr{D}_x & 0 & 0 \\ \mathscr{D}_z & 0 & 0 \end{bmatrix} \tag{40}
$$

so the adjoint equations according to Eq. (2) become

$$
\begin{aligned}
\frac{1}{\rho c^2}\frac{\partial \lambda_1}{\partial t} &= -\frac{\partial\left(\frac{1}{\rho}\lambda_2\right)}{\partial x} - \frac{\partial\left(\frac{1}{\rho}\lambda_3\right)}{\partial z} \\
\frac{\partial \lambda_2}{\partial t} &= -\frac{\partial \lambda_1}{\partial x} \\
\frac{\partial \lambda_3}{\partial t} &= -\frac{\partial \lambda_1}{\partial z}
\end{aligned} \tag{41}
$$

and the gradients according to Eq. (3)

$$
\begin{aligned}
\frac{d\chi}{dc} &= -\frac{2}{\rho c^3}\int_0^T \vec{\lambda_1}^T \vec{p}\, dt \\
\frac{d\chi}{d\rho} &= \int_0^T \frac{-1}{\rho^2 c^2}\vec{\lambda_1}^T \vec{p} + \frac{1}{\rho^2}\left(\vec{\lambda_2}^T \frac{\partial \vec{p}}{\partial x} + \vec{\lambda_3}^T \frac{\partial \vec{p}}{\partial z}\right)\, dt
\end{aligned} \tag{42}
$$

References

[1] Albert Tarantola. A strategy for nonlinear elastic inversion of seismic reflection data. *GEOPHYSICS*, 51(10):1893–1903, 1986.

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

[3] Andrew M. Bradley. PDE-constrained optimization problems and the adjoint method. 2012(2):1–6, 2012.

[4] Jean Virieux. SH-wave propagation in heterogeneous media : Velocity-stress finite-difference method. *Geophysics*, 49(11):1933–1957, 1984.

[5] Dimitri Komatitsch and Roland Martin. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics*, 72(5):SM155–SM167, 2007.

[6] Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves, 1994.

[7] Thomas M. Brocher. Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America*, 95(6):2081–2092, 2005.