Karlsruhe Institute of Technology

SEISMIC FULL WAVEFORM INVERSION

Lecture notes written by

VALERIE KRAMPE (NOW AT ETH ZÜRICH)

THOMAS BOHLEN

KIT - Faculty of Physics
Geophysical Institute (GPI)

WS 2021/22

Contents

1. Introduction

Seismic full waveform inversion (FWI) aims at a complete seismic characterization by exploiting the full signal content of seismic measurements. The aim of FWI is to fit iteratively the relevant signals (amplitude and phase) by numerical solutions of the full wave equation. The main product of FWI are multi-parameter models of acoustic/elastic/viscoelastic/anisotropic material properties which can explain the traveltime and amplitudes of (selected) recorded signals. Such multi-parameter models can help to improve the petrophysical characterization of geological site properties in-situ. In combination with the complementary information from reflection seismic images, FWI can thus help to characterize and monitor the in-situ conditions of rocks.

A second fundamental advantage of FWI is the possibility of utilizing various wave phenomena. In principle, all wave types predictable by the full wave equation can be exploited. This can include reflected waves, multiples, refracted waves, guided waves, mode conversions between P- and S-waves, (tunnel) surfaces waves, borehole guided waves etc. Unconventional wavefields often appear in complex environments in the presence of strong material discontinuities. Although each wave phenomena will require a special FWI workflow, they can be exploited to image specific features which conventional waves don't "see".

In recent 20 years FWI has received great attention and has been applied sucessfully to a broad range of spatial scales and wave types (Figure 1.1).

In a nutshell FWI is a PDE constrained local optimization procedure to iteratively fit observed data. The widely applied FWI workflow is illustrated in Figure 1.2. Starting with an initial model we iteratively calculate model updates utilizing the gradient of the misfit function which can be effectively calculated by the adjoint-state method. Without processing gradient based FWI would converge to the next local minimum of the misfit function which should be avoided. Most strategies change the topology of the misfit function during iterations by various means such as windowing the data in time and frequency, smoothing the model or the gradient, or changing the misfit function itself during iterations. The most important technique is called multi-scale approach which uses low frequencies (long wavelength) first and the gradually increases the bandwidth of the data to resolve finer details making the inverse problem gradually more non-linear. The features of the FWI workflow must be adapted to the problem under investigation. In marine environment FWI is often performed using the acoustic approximation as the wavefield is dominated by compressional waves. Recent applications of FWI to marine data have been very impressive. The resolution of the retrieved P-wave velocity models could be increased significantly compared to traveltime tomography reconstructions (**?**). In contrast, applications to land seismic data are much more challenging. The reasons are (1) the higher complexity of land seismic data due to the presence of P-, S-, and surface waves and (2) current limitations of the FWI technology in recovering multi-parameter models in viscoelastic media.

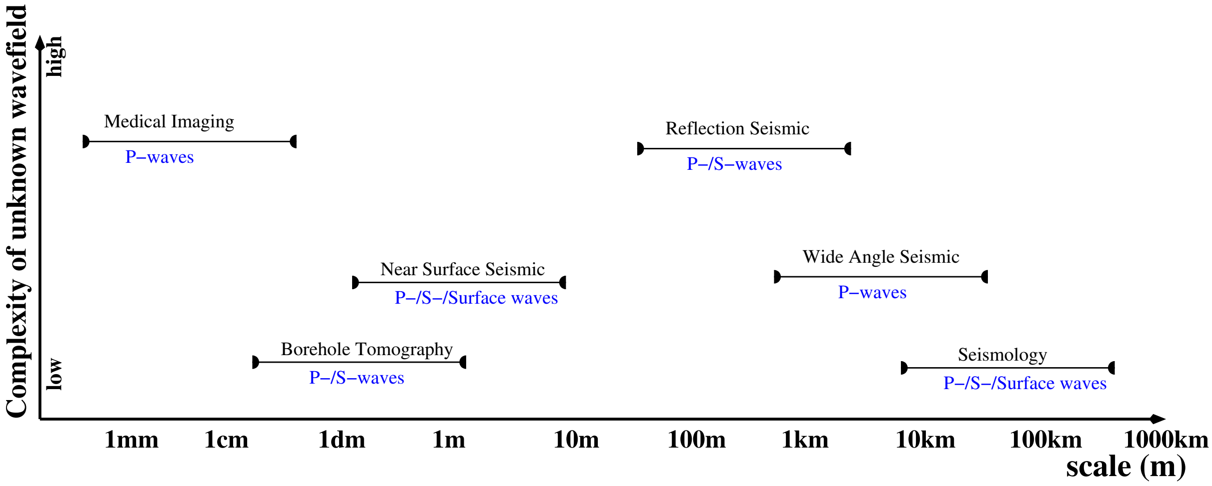


Figure 1.1.: *Today, 2D acoustic/elastic FWI has become feasible on a wide range of applications/scales covering nine orders of magnitude. Different wave types and acquisition configurations (transmission versus reflection) are applied. Applications in which the wavefield is composed of multiple scattered/reflected waves which is the case in medical imaging and reflection seismics are most challenging.*

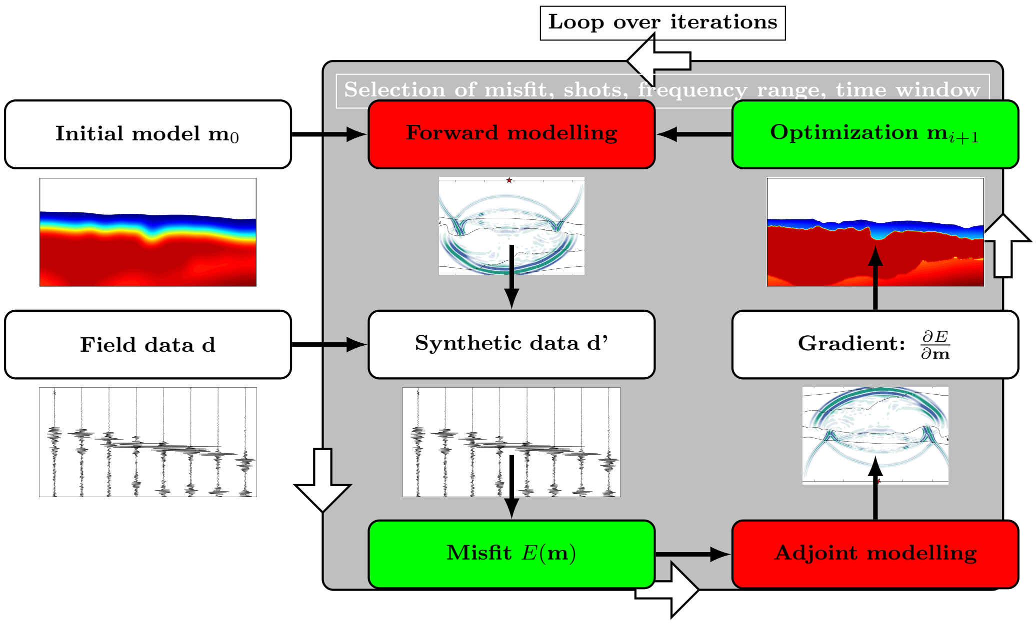


Figure 1.2.: *A typical workflow of gradient-based FWI. An initial model that already predicts the (time windowed) relevant signals at the lowest frequencies within half a cycle is required. For this model the synthetic seismic response for one or several shots is calculated and compared with the observed signals. Different misfit functions can be defined at different iterations to quantify the difference between observed and synthetic signals in an appropriate way. The misfit definition is crucial and drives the "adjoint" modelling where the corresponding adjoint wave equations are solved. The correlation of forward and adjoint wavefield provide the gradients for all model parameters simultaneously. This gradient gives the direction to update the model to reduce the chosen misfit. Higher order optimization methods that employ second order derivatives can improve the convergence and reduce parameter trade offs. The procedure is repeated for other selected shots, higher frequencies, other signals until the procedure has reached a sufficient local minimum of misfits. The white boxes indicate task may require large storage of wavefield in space and time. The red boxes indicate the modelling tasks which may demand high computing time. The green boxes indicate the "critical" choices of misfit function and optimization procedure that mainly steer the convergence towards the local minimum.*

2. Calculation of gradient

2.1. Born approximation

2.1.1. Scattering series

For an inversion, it would be ideal if there would be a linear relation between the model space $m(x)$ and the data space $u(x)$. To get such a linear relation, we develop a perturbation series, based on a Taylor series. The first order of this series is called the *Born approximation*. It is a fundamental relation between the data and model spaces and forms the basis for most migrations (e.g. reverse time migration) and full waveform inversion methods.

To develop the scattering series, we have to introduce perturbations for both the model and data space. As model parameters, we define

$$m(x) := \frac{1}{c^2(x)}, \tag{2.1}$$

and split it up into

$$m(x) = m_0(x) + \varepsilon m_1(x), \tag{2.2}$$

where $m_0(x)$ is the background model and $\varepsilon m_1(x)$ a small perturbation of this model.
We do the same for the data $u(x)$, splitting it up into

$$u(x) = u_0(x) + u_{sc}(x), \tag{2.3}$$

where $u_0(x)$ is the wavefield that would result if the model was only the background model m_0, and $u_{sc}(x)$ contains the perturbations of this wavefield caused by the perturbations of the velocity model. It can be interpreted as a result of scattering at the model perturbations, therefore it will here be called the scattered wavefield u_{sc}.

Because the background wavefield u_0 solves the wave equation (A.1) for the undisturbed model m_0, we get two wave equations, one for the background wavefield and one for the total wavefield:

$$m_0 \frac{\partial^2 u_0}{\partial t^2} - \Delta u_0 = f(x,t) \qquad \text{background wavefield} \tag{2.4}$$

$$m \frac{\partial^2 u}{\partial t^2} - \Delta u = f(x,t) \qquad \text{total wavefield} \tag{2.5}$$

Subtracting equation (2.4) from (2.5) yields

$$\begin{aligned}
& m\frac{\partial^2 u}{\partial t^2} - m_0\frac{\partial^2 u_0}{\partial t^2} - \Delta u + \Delta u_0 &=& \ 0 \\
\Leftrightarrow \quad & (m_0 + \varepsilon m_1)\frac{\partial^2 u}{\partial t^2} - m_0\frac{\partial^2}{\partial t^2}(u - u_{sc}) - \Delta(u - u_0) &=& \ 0 \\
\Leftrightarrow \quad & m_0\frac{\partial^2 u_{sc}}{\partial t^2} - \Delta u_{sc} &=& \ -\varepsilon m_1 \frac{\partial^2 u}{\partial t^2} \qquad (2.6)
\end{aligned}$$

This is again a wave equation for u_{sc} with the term $-\varepsilon m_1 \frac{\partial^2 u}{\partial t^2}$ as source. The scattered wavefield is thus generated by scattering of the total wavefield at the model perturbations m_1.

The solution of equation (2.6) can be found by convolving the Green's function with the source term (see equation (A.3)):

$$u_{sc}(x,t) = \int_0^t \int_{\mathbb{R}^3} G_0(x,y,t-s)(-\varepsilon m_1 \frac{\partial^2 u}{\partial t^2}(y,s))\ dy\ ds \tag{2.7}$$

It has to be considered, that the Green's function depends on the respective model. In this case, the Green's function G_0 for the background model m_0 is used.

In the following, we will use a shorter notation for the space-time integral by introducing an operator $\hat{G}_0$ that replaces the convolution and the space integral. Thus, equation (2.7) is written as

$$u_{sc} = -\varepsilon \hat{G}_0 m_1 \frac{\partial^2 u}{\partial t^2} = u - u_0 \tag{2.8}$$

which gives us the solution for the total wavefield

$$u = u_0 - \varepsilon \hat{G}_0 m_1 \frac{\partial^2 u}{\partial t^2} \tag{2.9}$$

This equation gives an implicit relation for the wavefield u and is called the *Lippmann-Schwinger equation*.

If we rearrange equation 2.9 and use a notation with operators, we get

$$\begin{aligned} \left[\hat{I} + \varepsilon \hat{G}_0 m_1 \frac{\partial^2}{\partial t^2}\right] u \ &= \ u_0 \\ \Leftrightarrow \qquad u \ &= \ \underbrace{\left[\hat{I} + \varepsilon \hat{G}_0 m_1 \frac{\partial^2}{\partial t^2}\right]^{-1}}_{\text{scattering operator}} \ u_0 \end{aligned} \tag{2.10}$$

with the identity operator $\hat{I}$. The scattering operator describes the relation between the background field u_0 and the total field u.

For an operator $\hat{A}$ we can develop the expression $[\hat{I} + \hat{A}]^{-1}$ in a Neumann series (?):

$$[\hat{I} + \hat{A}]^{-1} = \hat{I} - \hat{A} + \hat{A}^2 - \hat{A}^3 + ... \tag{2.11}$$

With $\hat{A} = \varepsilon \hat{G}_0 m_1 \frac{\partial^2}{\partial t^2}$, equation (2.10) can be written as a series, which is called the *Born series*:

$$\begin{aligned} u &= u_0 - \varepsilon(\hat{G}_0 m_1 \frac{\partial^2}{\partial t^2})u_0 + \varepsilon^2 (\hat{G}_0 m_1 \frac{\partial^2}{\partial t^2})(\hat{G}_0 m_1 \frac{\partial^2}{\partial t^2})u_0 + ... \\ &= u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + ... \end{aligned} \tag{2.12}$$

The first order term $u_1 = -\hat{G}_0 m_1 \frac{\partial^2 u_0}{\partial t^2}$, corresponds to single scattering, while the second order term $u_2 = (\hat{G}_0 m_1 \frac{\partial^2}{\partial t^2})(\hat{G}_0 m_1 \frac{\partial^2}{\partial t^2})u_0$ corresponds to double scattering, meaning that one wave is scattered at two different points on its travel path.

The first order approximation $u = u_0 + \varepsilon u_1$, that is considering only single scattering, is called the

Born approximation and will be used in the following. With this approximation, the scattered wavefield is given by $u_{sc} = \varepsilon u_1$. Inserting this in equation (2.6) results in

$$m_0 \frac{\partial^2 \varepsilon u_1}{\partial t^2} - \Delta \varepsilon u_1 = -\varepsilon m_1 \frac{\partial^2 (u_0 + \varepsilon u_1)}{\partial t^2}$$
$$\Leftrightarrow \quad \boxed{m_0 \frac{\partial^2 u_1}{\partial t^2} - \Delta u_1 = -m_1 \frac{\partial^2 u_0}{\partial t^2}} \tag{2.13}$$

where we neglected the term of the order ε^2.
This is now a wave equation for the single scattered field. In the source term appears the background wavefield u_0, so the scattered wavefield is induced by a secondary source that corresponds to the background wavefield scattered at the model perturbations m_1. This secondary scattered wavefield is then propagating in the background model m_0. The background wavefield u_0 can be calculated as before with equation (2.4).
Figure 2.1 shows the principle of the scattering theory: The source $f(x,t)$ is initiating the background wavefield u_0. At model perturbations εm_1, the background field is scattered, which leads to the additional wavefield u_1. Because we have different positions of sources and receivers, the Green's function differs for both parts of the wavefield. For the background field, we use the notation $\hat{G}_0 = G(y, x_s, t)$, and for the scattered field with the secondary source $\hat{G}'_0 = G(x_r, y, t)$. Thus, the wavefields are calculated with the Green's functions as follows:

$$u_0 = \hat{G}_0 f \tag{2.14}$$

$$u_1 = -\hat{G}'_0 m_1 \frac{\partial^2 u_0}{\partial t^2} = -m_1 \hat{G}'_0 \frac{\partial^2}{\partial t^2}(\hat{G}_0 f) \tag{2.15}$$

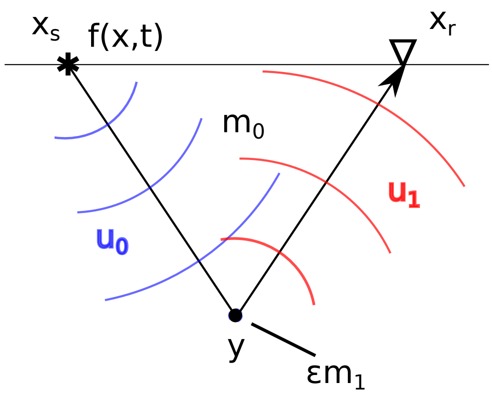


Figure 2.1.: Principle of the scattering theory: The from the source f induced background wavefield u_0 is scattered at the model perturbations εm_1 in point y, producing the scattered wavefield u_1

With equation (2.15), we finally have a linear relation between the scattered wavefield u_1 and the model perturbations m_1. For this linear function, it is easy to calculate the derivative of the wavefield u_1 with respect to the model parameters m_1, which is called the *Frechet-derivative*:

$$\frac{\partial u_1}{\partial m_1} = -\hat{G}'_0 \frac{\partial^2}{\partial t^2}(\hat{G}_0 f) \tag{2.16}$$

If we transform the Frechet-derivative into the frequency domain, the second partial derivative $\frac{\partial^2}{\partial t^2}$ will be replaced by $-w^2$, so that we get

$$\frac{\partial u_1}{\partial m_1} = w^2 \hat{\mathcal{G}}'_0 \hat{\mathcal{G}}_0 f, \tag{2.17}$$

where $\hat{\mathcal{G}}_0$ denotes the Fourier transform of the operator $\hat{G}_0$. This derivative will be used for the inversion as described in the following section.

2.2. Inversion of the scattered wavefield

Now that we have linearized the relation between the scattered field u_1 and the model perturbations m_1, we can invert this first order scattered wavefield. Therefore, we will use the steepest gradient method in the frequency domain.

First, we formulate the L2-misfit function E that should be minimized in the inversion. In frequency domain, the misfit is

$$E = \frac{1}{2} \int_w |u_1|^2 \, dw \tag{2.18}$$

This means, that the inversion should find a model, for which in the ideal case there is no scattered wavefield. Then, the true model would be the "background" model m_0 and there would be no further model perturbations. With real data, the ideal case will not be reached, but only a minimum of the misfit function.

The gradient descent method is an iterative inversion method. In each iteration step, the gradient of the misfit function with respect to the model parameters $\frac{\partial E}{\partial m}$ is calculated, and the model parameters are changed along the direction of this gradient:

$$m^{(n+1)} = m^{(n)} - \alpha \frac{\partial E}{\partial m} \tag{2.19}$$

By this way, it is assured that the misfit is getting smaller in the next step. Thus, finally a model with a minimal misfit will be reached. The parameter α in equation (2.19) is the step length and defines, how much the model parameters can change in one iteration step.

For the calculation of the misfit gradient, we split the wavefield u_1 (which is complex in the frequency domain) up into its real part $u_{1,R}$ and its imaginary part $u_{1,I}$ and insert it in equation (2.18):

$$\frac{\partial E}{\partial m} = \frac{1}{2} \frac{\partial}{\partial m} \int_w |u_{1,R} + i u_{1,I}|^2 \, dw \tag{2.20}$$

The partial derivative of the complex function is calculated as follows:

$$\begin{aligned} \frac{\partial}{\partial m} |u_{1,R} + i u_{1,I}|^2 &= \frac{\partial}{\partial m} (u_{1,R}^2 + u_{1,I}^2) = 2 \left(u_{1,R} \frac{\partial u_{1,R}}{\partial m} + u_{1,I} \frac{\partial u_{1,I}}{\partial m} \right) \\ &= 2 \mathrm{Re} \left[\left(\frac{\partial u_{1,R}}{\partial m} + i \frac{\partial u_{1,I}}{\partial m} \right) (u_{1,R} - i u_{1,I}) \right] \\ &= 2 \mathrm{Re} \left[\frac{\partial u_1}{\partial m} u_1^* \right] \end{aligned}$$

Inserting this in equation (2.20) yields

$$\frac{\partial E}{\partial m} = \int_w \mathrm{Re} \left[\frac{\partial u_1}{\partial m} u_1^* \right] \, dw \tag{2.21}$$

We now use the Born approximation for the Frechet-derivative ($\frac{\partial u_1}{\partial m} = w^2 \hat{\mathcal{G}}_0' \hat{\mathcal{G}}_0 f$, see equation (2.17)), leading to the misfit gradient

$$\boxed{\frac{\partial E}{\partial m} = \int_w w^2 \ \mathrm{Re}\left[\hat{\mathcal{G}}_0 f \ \hat{\mathcal{G}}_0' u_1^*\right] \ dw} \tag{2.22}$$

In this equation, $\hat{\mathcal{G}}_0 f$ can be physically interpreted as the forward-propagating wave induced by the source f, and $\hat{\mathcal{G}}_0' u_1^*$ as the scattered field u_1 propagating from the receiver backward in time, because the complex conjugation in frequency domain corresponds to a reversion of time in time domain.

Thus, the misfit gradient in frequency domain can be calculated with the following four steps:

1) Calculate the forward wavefield $\hat{\mathcal{G}}_0 f$ from the source

2) Backpropagate the scattered wavefield $\hat{\mathcal{G}}_0' u_1^*$ from the receiver

3) Multiply the wavefields

4) Sum over all frequencies

In FWI, we set $u_1 = u_{obs} - u_{synth}$, where u_{obs} is the measured data and u_{synth} the data that is resulting from the currently used model. u_1 is then the missing/residual wavefield and is interpreted as the single scattered wavefield in terms of the Born approximation.

For the calculation of the gradient, we still have the problem that we need to know the Green's functions, which becomes impossible for complex models. To resolve this problem, we will introduce in the following sections the linearized forward and the adjoint operator, which will enable us to calculate the misfit gradient in time domain.

3. The adjoint-state method

3.1. Adjoint method

3.1.1. Linearization of the forward operator

In this section, we present a second method to derive the linear relation between data and model that resulted from the Born approximation in section 3.2.1
Again, we split our wavefield $u = u_0 + u_1$ up in background and scattered field. Our model parameters are $m = m_0 + m_1$, where m_1 describes the perturbations of the background model (Note: We changed the notation compared to the sections before, no longer using ε). We introduce the forward operator $\tilde{F}$, that describes the relation between model m and data u, so $u = \tilde{F}[m]$. The Taylor series of u near to $m = m_0$ then is

$$u = u_0 + \frac{\partial \tilde{F}}{\partial m}[m_0]m_1 + \frac{1}{2}\langle \frac{\partial^2 \tilde{F}}{\partial m^2}[m_0]m_1, m_2\rangle \tag{3.1}$$

We will consider only the first order term of this Taylor series, because the following terms describe the wavefield that is generated by double, triple, etc. scattering. As for the Born approximation, we only consider single scattering.
We define the linear operator

$$F := \frac{\partial \tilde{F}}{\partial m} = \frac{\partial u}{\partial m} \tag{3.2}$$

Compared with equation (3.1), this yields

$$u = u_0 + Fm_1 \qquad \Rightarrow \qquad u_1 = Fm_1, \tag{3.3}$$

so that F describes a linear relation between u_1 and m_1.

To derive with this the Born approximation, we differentiate the acoustic wave equation (A.1) with respect to m, which gives

$$\frac{\partial^2 u}{\partial t^2} + m\frac{\partial^2}{\partial t^2}\frac{\partial u}{\partial m} - \Delta\frac{\partial u}{\partial m} = 0 \tag{3.4}$$

Using (3.2) and (3.3), we replace $\frac{\partial u}{\partial m} = F = \frac{u_1}{m_1}$, resulting in

$$\begin{aligned} &\frac{\partial^2 u}{\partial t^2} + (m_0 + m_1)\frac{\partial^2}{\partial t^2}\frac{u_1}{m_1} - \Delta\frac{u_1}{m_1} = 0 \\ \Leftrightarrow \qquad &m_1\frac{\partial^2}{\partial t^2}(u_0 + u_1) + (m_0 + m_1)\frac{\partial^2 u_1}{\partial t^2} - \Delta u_1 = 0 \end{aligned}$$

In this equation, we can neglect the term $m_1\frac{\partial^2 u_1}{\partial t^2}$, because those terms of higher order are describing multiple scattering, that we do not consider in this approximation. When neglecting those terms, we get the wave equation

$$m_0\frac{\partial^2 u_1}{\partial t^2} - \Delta u_1 = -m_1\frac{\partial^2 u_0}{\partial t^2}, \tag{3.5}$$

which is the same than equation (2.13) that resulted from the Born scattering theory. The solution for the scattered field u_1 is then given by the Green's function as

$$u_1 = -\hat{G}_0 m_1 \frac{\partial^2 u_0}{\partial t^2} \tag{3.6}$$

3.1.2. Adjoint operator of the linear operator F

We first introduce the needed mathematical theory concerning adjoint operators. For further details, see **?**.

If H_1 and H_2 are Hilbert spaces and the operator $T : \mathrm{H}_1 \to \mathrm{H}_2$ is linear, then the adjoint linear operator $T^* : \mathrm{H}_2 \to \mathrm{H}_1$ can be found with $x \in \mathrm{H}_1$ and $y \in \mathrm{H}_2$ via

$$\langle Tx, y \rangle_{H_2} = \langle x, T^* y \rangle_{H_1}, \tag{3.7}$$

where $\langle\rangle_{H_1}$ and $\langle\rangle_{H_2}$ denote scalar products in H_1 and H_2, respectively. In the integral notation of the scalar products, this means

$$\int_{H_2} Tx(r)\ y(r)\ dr_{H_2} = \int_{H_1} x(r)\ T^* y(r)\ dr_{H_1} \tag{3.8}$$

We now apply this on our problem $u_1 = Fm_1$. The adjoint operator F^* thus fulfills the equation

$$\langle d_{obs}, \underbrace{Fm_1}_{u_1} \rangle = \langle F^* d_{obs}, m_1 \rangle, \tag{3.9}$$

where the left side is a scalar product in data space and the right side one in model space. The adjoint operator is thus a linear map from data into model space, while F is a linear map from model to data space.
Thus, we can interpret F as forward operator and F^* as *imaging* or *backward* operator, which projects the observed data back to the model parameters that caused the observed wavefield (see figure 3.1).

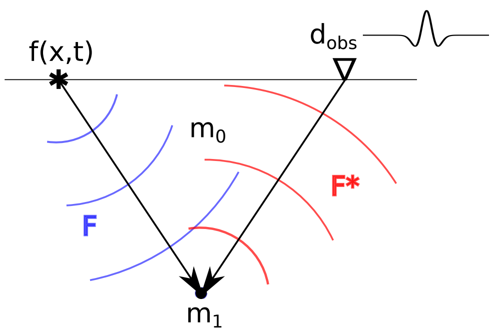


Figure 3.1.: Interpretation of the forward operator F and its adjoint operator F^*, that works as a backward propagator

In FWI, the adjoint operator is of importance in the context of minimization. The goal of the inversion is to minimize the L2 misfit function

$$
\begin{aligned}
J(m) &= \frac{1}{2}||d_{obs} - \tilde{F}[m]||_2 \\
&= \frac{1}{2}\sum_{r,s}\int_0^T |d_{obs} - \tilde{F}[m]|^2 \, dt,
\end{aligned} \tag{3.10}
$$

where $d_{obs} - \tilde{F}[m]$ is the difference between the observed data and the data that results from the model, and $\sum_{r,s}$ denotes the summation over all sources and receivers.
We will show in the following that the misfit gradient $\frac{\partial J}{\partial m}$ can be expressed via

$$
\boxed{\frac{\partial J}{\partial m} = F^*(\tilde{F}[m] - d_{obs})} \tag{3.11}
$$

With this, we can use the steepest gradient method to update the model in each iteration step via

$$
m^{(n+1)} = m^n - \alpha\frac{\partial J}{\partial m} \tag{3.12}
$$

This method is called the *adoint method*, because the adjoint operator F^* is used to calculate the misfit gradient.

If we know F^*, we can thus perform

- Imaging by applying F^* one time to the observed data (Reverse Time Migration)
- Full waveform inversion by applying F^* iteratively to the residual wavefield $\tilde{F}[m] - d_{obs}$

We now want to proof equation (3.11). The linearized forward operator F is defined as in section 3.3.1 as $F = \frac{\partial \tilde{F}}{\partial m}[m]$. With that, the Taylor expansion of $\tilde{F}$ gives

$$
\tilde{F}[m+h] = \tilde{F}[m] + Fh + O(h^2) \tag{3.13}
$$

Inserting this into the misfit function $J(m+h)$ yields

$$
\begin{aligned}
J(m+h) &= \frac{1}{2}\langle \tilde{F}[m+h] - d_{obs}, \tilde{F}[m+h] - d_{obs}\rangle \\
&= \frac{1}{2}\langle \tilde{F}[m] - d_{obs}, \tilde{F}[m] - d_{obs}\rangle + \langle Fh, \tilde{F}[m] - d_{obs}\rangle + O(h^2) \\
&= J(m) + \langle h, F^*(\tilde{F}[m] - d_{obs})\rangle + O(h^2)
\end{aligned} \tag{3.14}
$$

This corresponds to the Taylor series of J(m+h). The comparison of equation (3.14) with the general Taylor expansion

$$
J(m+h) = J(m) + \langle h, \frac{\partial J}{\partial m}\rangle + O(h^2) \tag{3.15}
$$

gives us as result for the misfit gradient

$$
\frac{\partial J}{\partial m} = F^*(\tilde{F}[m] - d_{obs}), \tag{3.16}
$$

which is exactly the equation of the adjoint method that we wanted to proof.

In this equation, the expression on which the adjoint operator is applied is called the *adjoint* *source*. It depends on the misfit function that is used. For the L2 norm as misfit function, the adjoint source is the residual wavefield $\tilde{F}[m] - d_{obs}$.

3.1.3. Calculation of F^* for the wave equation

As we showed in the previous section, it is necessary for the FWI to know the adjoint operator F^*. In this section, it will be derived for the wave equation.

The starting point is the equality of scalar products in the two domains (see equation (3.7)):

$$\underbrace{\langle d_{obs}, Fm_1 \rangle}_{\text{data space}} = \underbrace{\langle F^* d_{obs}, m_1 \rangle}_{\text{model space}} \tag{3.17}$$

For the moment, we consider only one source but multiple receivers at the locations x_r, so that

$$d_{obs}(x,t) = \sum_r d_r(t)\delta(x - x_r) \tag{3.18}$$

We will come back to the multiple-source problem in section 4.1.

With $Fm_1 = u_1$, equation (3.17) can be written as

$$\int_{\mathbb{R}^3} \int_0^T d_{obs}(x,t)\ u_1(x,t)\ dt\ dx = \int_{\mathbb{R}^3} F^* d_{obs}(x)\ m_1(x)\ dx \tag{3.19}$$

The relation $u_1 = Fm_1$ implies the two wave equations

$$(m_0 \frac{\partial^2}{\partial t^2} - \Delta)u_0 = f \tag{3.20}$$

$$(m_0 \frac{\partial^2}{\partial t^2} - \Delta)u_1 = -m_1 \frac{\partial^2 u_0}{\partial t^2}, \tag{3.21}$$

which are describing the generation and propagation of the background wavefield u_0 and the scattered wavefield u_1, respectively (see section 3.3.1).

We now consider an auxiliary wavefield $q(x,t)$, the *adjoint wavefield*, that solves the wave equation with d_{obs} as right-hand side:

$$(m_0 \frac{\partial^2}{\partial t^2} - \Delta)q(x,t) = d_{obs}(x,t) \tag{3.22}$$

It can be interpreted as wavefield that is propagating backward in time with the observed data as source.
We substitute this into $\langle d_{obs}, Fm_1 \rangle$ and obtain

$$\begin{aligned} \langle d_{obs}, Fm_1 \rangle &= \int_{\mathbb{R}^3} \int_0^T (m_0 \frac{\partial^2}{\partial t^2} - \Delta) q(x,t) u_1(x,t)\ dt\ dx \\ &= \int_{\mathbb{R}^3} \int_0^T m_0 \frac{\partial^2 q}{\partial t^2} u_1\ dt\ dx - \int_{\mathbb{R}^3} \int_0^T \Delta q\ u_1\ dt\ dx \end{aligned} \tag{3.23}$$

We apply two times a partial integration in both space and time on this equation. The first integral is integrated in time, the second one in space. For the first one, we obtain

$$\begin{aligned}\int_{\mathbb{R}^3}\int_0^T m_0\frac{\partial^2 q}{\partial t^2}u_1\ dt\ dx &= \int_{\mathbb{R}^3} m_0\frac{\partial q}{\partial t}u_1\bigg|_0^T\ dx - \int_{\mathbb{R}^3}\int_0^T m_0\frac{\partial q}{\partial t}\frac{\partial u_1}{\partial t}\ dt\ dx\\ &= \int_{\mathbb{R}^3} m_0\frac{\partial q}{\partial t}u_1\bigg|_0^T\ dx - \int_{\mathbb{R}^3} m_0 q\frac{\partial u_1}{\partial t}\bigg|_0^T\ dx + \int_{\mathbb{R}^3}\int_0^T m_0 q\frac{\partial^2 u_1}{\partial t^2}\ dt\ dx\end{aligned} \tag{3.24}$$

For the second one, we use the theorem of Green, which transforms a volume integral into a surface integral via

$$\int_V (u_1\Delta q - q\Delta u_1)\ dx = \int_{\partial V}\left(u_1\frac{\partial q}{\partial n} - q\frac{\partial u_1}{\partial n}\right)\ dS_x, \tag{3.25}$$

where ∂V is the surface of the volume V, $\frac{\partial q}{\partial n}$ is the derivative in direction of the normal vector of the surface and dS_x means integration over the whole surface (?).
With this theorem, the second integral in equation (3.23) can be written as

$$\int_V\int_0^T \Delta q\ u_1\ dt\ dx = \int_V\int_0^T q\Delta u_1\ dt\ dx + \int_{\partial V}\int_0^T u_1\frac{\partial q}{\partial n}\ dS_x\ dt - \int_{\partial V}\int_0^T q\frac{\partial u_1}{\partial n}\ dS_x\ dt, \tag{3.26}$$

where we substituted the space integration over $\mathbb{R}^3$ with an finite volume V. This makes no difference, because V can be chosen so big that the integration would be almost the same than for the infinite space.

Together with (3.24), the result for the scalar product is

$$\begin{aligned}\langle d_{obs}, Fm_1\rangle = &\int_V\int_0^T q\left(m_0\frac{\partial^2}{\partial t^2} - \Delta\right)u_1\ dx\ dt + \int_V m_0\frac{\partial q}{\partial t}u_1\bigg|_0^T\ dx - \int_V m_0 q\frac{\partial u_1}{\partial t}\bigg|_0^T\ dx\\ &- \int_{\partial V}\int_0^T u_1\frac{\partial q}{\partial n}\ dS_x\ dt + \int_{\partial V}\int_0^T q\frac{\partial u_1}{\partial n}\ dS_x\ dt\end{aligned} \tag{3.27}$$

This can be simplified significantly by introducing boundary conditions. Because we chose the volume V so large that it covers nearly the space $\mathbb{R}^3$, it is impossible for the waves to reach the boundary of this volume in the considered travel time T. Thus, the integrations over the boundary of V can be neglected. Additionally, we use the starting conditions $u(t=0)=0$ and $\frac{\partial u}{\partial t}(t=0)=0$ and the final conditions $q(t=T)=0$ and $\frac{\partial q}{\partial t}(t=T)=0$. Then, equation (3.27) is reduced to

$$\begin{aligned}\langle d_{obs}, Fm_1\rangle &= \int_V\int_0^T q(x,t)\left(m_0\frac{\partial^2}{\partial t^2} - \Delta\right)u_1(x,t)\ dx\ dt\\ &= -\int_V\int_0^T q(x,t)m_1\frac{\partial^2 u_0}{\partial t^2}\ dx\ dt,\end{aligned} \tag{3.28}$$

where we used equation (3.21) for the last step.
Because of $\langle d_{obs}, Fm_1\rangle = \langle F^* d_{obs}, m_1\rangle$, this gives us the result

$$\boxed{F^* d_{obs} = -\int_0^T q(x,t)\frac{\partial^2 u_0}{\partial t^2}(x,t)\ dt} \tag{3.29}$$

It corresponds to a zero-lag cross-correlation of the adjoint field $q(x,t)$ (backpropagating) and the second derivative of the forward propagating background field $u_0(x,t)$. In this form, it is the imaging condition for the Reverse Time migration. The adjoint field q corresponds here to the time reversed observed data.

If we now come back to the relation used in FWI

$$\frac{\partial J}{\partial m}(m) = F^*(\tilde{F}[m] - d_{obs}) \tag{3.30}$$

(see (3.11)), and compare it with the imaging condition (3.29), we conclude that we just have to modify the adjoint source in the adjoint equation (3.22)

$$(m_0 \frac{\partial^2}{\partial t^2} - \Delta) q(x,t) = f_{adj} \tag{3.31}$$

If we use the observed data as adjoint source, i.e. $f_{adj} = d_{obs}$, we get the imaging condition for the Reverse Time Migration (3.29). But if we use instead the residual wavefield $\tilde{F}[m] - d_{obs}$, we get the result

$$\boxed{F^*(\tilde{F}[m] - d_{obs}) = -\int_0^T q(x,t) \frac{\partial^2 u_0}{\partial t^2}(x,t)\ dt = \frac{\partial J}{\partial m}(m)}, \tag{3.32}$$

where $q(x,t)$ is now the backward propagated wavefield with the residuals of observed and synthetic data as source. This equation makes it possible to calculate the misfit gradient, that can then be used to update the model iteratively with the steepest gradient method.

4. Adaption of FWI for real data

4.1. Multiple sources

In the last section, we derived the imaging condition for the Reverse Time Migration for one source:

$$F^* d_{obs} = -\int_0^T q(x,t) \frac{\partial^2 u_0}{\partial t^2}(x,t)\ dt \tag{4.1}$$

For multiple sources and receivers, we have one observed time series for each pair of source and receiver, so we write the observed data as summation over those time series, i.e. $d_{obs} = \sum_s \sum_r d_{obs,r,s}$. We can then write the scalar product $\langle d_{obs}, Fm_1 \rangle$ as

$$\langle d_{obs}, Fm_1 \rangle = \langle d_{obs}, u_1 \rangle = \sum_s \sum_r \int_0^T d_{obs,r,s}(t) u_{1,s}(x_r,t)\ dt \tag{4.2}$$

Because it is no problem to calculate the wavefield for one source at many receivers, we can summarize the summation over all receivers for one source to $\langle d_{obs,s}, F_s m_1 \rangle$, so that

$$\langle d_{obs}, Fm_1 \rangle = \sum_s \langle d_{obs,s}, F_s m_1 \rangle \tag{4.3}$$

The adjoint operator is then calculated as follows:

$$\begin{aligned} \langle F^* d_{obs}, m_1 \rangle = \langle d_{obs}, Fm_1 \rangle &= \sum_s \langle d_{obs,s}, F_s m_1 \rangle = \sum_s \langle F_s^* d_{obs,s}, m_1 \rangle \\ &= \langle \sum_s F_s^* d_{obs,s}, m_1 \rangle \end{aligned} \tag{4.4}$$

This shows, that

$$\boxed{F^* d_{obs} = \sum_s F_s^* d_{obs,s}} \tag{4.5}$$

which means that the imaging condition for multiple sources is the sum of the imaging conditions for each single source. It is calculate by

$$F^* d_{obs} = -\sum_s \int_0^T q_s(x,t) \frac{\partial^2 u_{0,s}}{\partial t^2}(x,t)\ dt, \tag{4.6}$$

where $u_{0,s}$ is the forward field for the source s and q_s the adjoint field for the source s. The adjoint field is the backpropagated field with the observed data $d_{obs,s}$ as source. Therefore, all receivers act simultaneously as seismic sources for the adjoint field.

To calculate the adjoint operator F^* for FWI, we only have to replace the adjoint source d_{obs} by the residuals $\tilde{F}[m] - d_{obs}$. The summation over all sources remains the same so that the misfit gradient for multiple sources is calculated via

$$\frac{\partial J}{\partial m} = F^*(\tilde{F}[m] - d_{obs}) = \sum_s F_s^*(\tilde{F}_s[m] - d_{obs,s}) \tag{4.7}$$

4.2. General procedure of adjoint Full Waveform Inversion

In this section, we formulate a general step-by-step procedure for an adjoint FWI with multiple sources and receivers. The steps are listed up in the following. Arrows connecting steps symbolize the loops that have to be implemented: The first one is needed to summarize over all sources, and the second one realizes the iterative inversion (steps 1-8 are one single iteration step).

0. Create starting model m_0

1. Forward calculation for each source in the current background model m_0
$\rightarrow u_{0,s}(x,t) = \tilde{F}[m_0]$

2. Store $u_{0,s}(x,t)$

3. Solve the adjoint equation with adjoint sources $F[\tilde{m_0}] - d_{obs}$ (backward propagation of adjoint sources) $\rightarrow q_s(x,t)$

4. Apply zero-lag cross-correlation to calculate misfit gradient:

$$\frac{\partial J_s}{\partial m} = -\int_0^T q_s(x,t)\frac{\partial^2 u_{0,s}}{\partial t^2}\,dt$$

Repeat steps 1. - 4. for all sources

5. Sum over all sources:

$$\frac{\partial J}{\partial m} = \sum_s \frac{\partial J_s}{\partial m}$$

6. Calculate the step length α or use a constant value

7. Update model:

$$m_1 = m_0 - \alpha\frac{\partial J}{\partial m}$$

8. Set updated model as new starting model: $m_1 = m_0$

Go back to 1. and repeat steps 1. - 8. with new starting model

9. When any stopping criterion is reached $\rightarrow$ final model m_1

The calculation of the step length (step 6) can be done in the following way: For a few selected sources, the misfit J is calculated for different step lengths α. The misfit values are plotted against α and fitted by a parabolic function. The minimum of this parabolic fit is reached for the optimal step length that will then be used for the model update.

In each iteration step, we have one forward calculation for the background field u_0 and one for the adjoint field q. Thus, the total amount of forward calculations for N iteration steps is $2N$ plus eventually some calculations for the step length α.

4.3. Inversion of the source time function

If we are working with field data, we often have the problem that the source function is not exactly known. Because of attenuation effects, even a receiver close to the source does not recover the source function accurately. However, for the FWI, the source function is needed for the forward calculation of the background field. To resolve this problem, we do a so-called inversion of the source time function, which is realized by a deconvolution.

The observed data d_{obs} is generated by the unknown true source s_{true}. We try to approximate this true source with synthetic source s that generates by forward calculation the synthetic data u. The synthetic field of one source at one receiver at position x_r can be calculated with the Green's function of the model m_0 via a convolution:

$$u(x_r, t) = \int_0^T G_{m_0}(x_s, x_r, t - t')s(t')\ dt' \tag{4.8}$$

A convolution in time domain corresponds to a multiplication in frequency domain:

$$u(x_r, w) = G_{m_0}(x_s, x_r, w)s(w) \tag{4.9}$$

Because the multiplication is easier to handle than the convolution, the deconvolution will be done in the frequency domain.

The observed data is calculated in the same way, only that the Green's function of the true model and the true source are used, yielding

$$d_{obs}(x_r, w) = G_{m_{true}}(x_s, x_r, w)s_{true}(w) \tag{4.10}$$

Our goal is to find a linear filter $c(w)$ in the frequency domain so that

$$s_{true}(w) - c(w)s(w) = \min \tag{4.11}$$

We could then approximate the true source by the filtered synthetic one. The filter $c(w)$ is called *source wavelet correction filter*.

Because the true source is unknown, it is not possible to use the condition (4.11) in this form. We first have to make an approximation by assuming that the Green's function is equal for the true and for the synthetic model, i.e.

$$G := G_{m_0} = G_{m_t rue} \tag{4.12}$$

Then we can multiply the minimum condition with G, which gives

$$G(s_{true} - cs) = d_{obs} - cu = \min \tag{4.13}$$

We can thus apply the filter c to the synthetic seismograms u instead of applying it on the synthetic source. The optimal filter will reduce the residuals of observed and synthetic seismograms to a minimum. Although the assumption $G_{m_0} = G_{m_t rue}$ is very strong, it works for FWI.

For minimization, we formulate the difference in (4.13) as L2-norm:

$$J_s = \sum_{k=1}^{M} \int_w |d_{obs}(x_k, w) - c(w)u(x_k, w)|^2 \; dw + \underbrace{\varepsilon^2 \int_w |c(w)|^2 \; dw}_{damping} \tag{4.14}$$

The summation over k considers that we have multiple (M) receivers at the locations x_k. The second term is a damping term that will assure numerical stability by avoiding a division by zero.
For discrete frequencies $w_l = l\Delta w$, $l = 0, ..., N-1$, we can write

$$J_s = \left(\sum_{k=1}^{M} \sum_{l=0}^{N-1} |d_{obs}(x_k, w_l) - c(w_l)u(x_k, w_l)|^2 + \varepsilon^2 \sum_{l=0}^{N-1} |c(w)|^2 \right) \Delta w \tag{4.15}$$

In frequency domain, the filter and the wavefields are consisting of complex values. We write the filter components $c(w_l)$ as sum of real and imaginary part:

$$c(w_l) = c_l = c_{l,R} + i c_{l,I} \tag{4.16}$$

The function J_s is minimized when all partial derivatives are zero, so

$$\frac{\partial J_s}{\partial c_{l,R}} = 0 \qquad \text{and} \qquad \frac{\partial J_s}{\partial c_{l,I}} = 0 \tag{4.17}$$

If we split also the wavefields d_{obs} and u up into their real and imaginary parts, we can calculate both partial derivatives. The summation of both results gives us the result for the filter components

$$\boxed{c_l = \frac{\sum_{k=1}^{M} u^*(x_k, w_l) d_{obs}(x_k, w_l)}{\varepsilon^2 + \sum_{k=1}^{M} |u(x_k, w_l)|^2}} \tag{4.18}$$

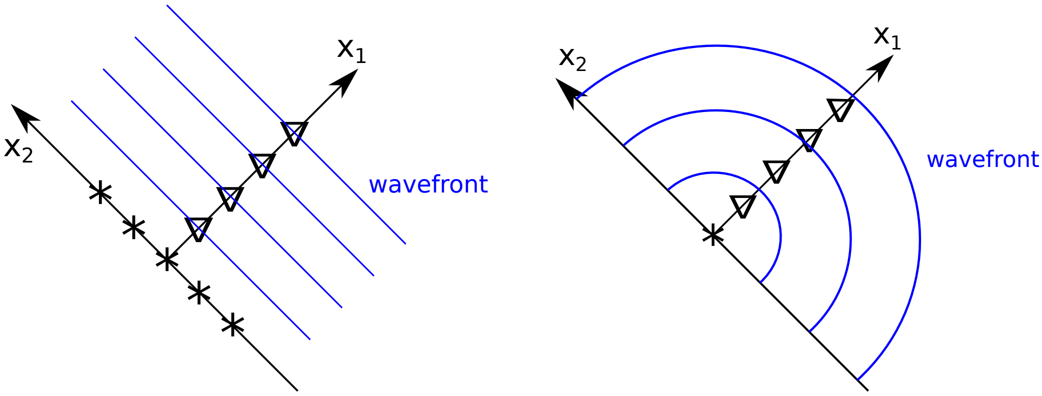
This filter is called *Wiener-filter* or *water-level deconvolution*.

For the application of the source time function inversion in FWI, there are some additional remarks:

- An advantage of this method is that the filter coefficients c_l are resulting from a direct inversion, i.e. only one iteration step is needed
- The filtered signal $c(w)s(w)$ corresponds approximately to the true source signal
- A stable and causal result for s_{true} indicates a stable convergence
- For the synthetic source s, any signal is possible, e.g. a δ-impulse or a Ricker wavelet
- In FWI, the source time function inversion is applied once per frequency interval
- The same method can be applied to invert receiver-function correction filters

4.4. Correction of geometrical spreading

If we want to apply a 2D FWI on 3D field data, it is necessary to correct the geometrical spreading, because the spreading in the 2D simulation differs from the real one. Field data is usually generated by a point source, e.g. by a hammer blow or an explosion. From this source, spherical waves are propagating. On the contrary, a 2D simulation implies a line source along the third (missing) dimension, which results in a propagation of plane waves (see figure 4.1).



(a) 2D simulation: Line source (b) 3D field data: Point source

Figure 4.1.: Differences of geometrical spreading between the 2D simulation (a)) and the 3D field data (b)). In 2D, a line source is assumed, while in 3D, we have a point source. x_2 is the third dimension that will be missing in the 2D simulation.

Our goal is to find a filter $F(r,k)$ in frequency domain that transforms the 3D data into 2D data assuming a line source. Here, r denotes the distance to the source. In order to find this filter, we consider the acoustic wave equation

$$\left[\frac{\partial^2}{\partial t^2} - c^2(x)\Delta\right] u(x,t) = f(x,t) \tag{4.19}$$

The Green's function $G(x, x_s, t)$ is the solution of

$$\left[\frac{\partial^2}{\partial t^2} - c^2(x)\Delta\right] G(x, x_s, t) = \delta(x - x_s)\delta t \tag{4.20}$$

with the source location x_s. In frequency domain, this equation corresponds to

$$\left[k^2 + \Delta\right] \mathcal{G}(x, x_s, w) = -4\pi\delta(x - x_s) \tag{4.21}$$

Assuming a constant velocity c, we obtain the follwing solutions in the far field:

$$\text{3D}: \qquad G^{3D}(x, x_s, w) = \frac{e^{ikr}}{r} \tag{4.22}$$

$$\text{2D}: \qquad G^{2D}(x, x_s, w) = \sqrt{\frac{2\pi}{kr}} e^{ikr} e^{i\pi/4} \tag{4.23}$$

with $r = ||x - x_s||$. These solutions can be proved by transforming the Laplace operator in equation (4.21) into spherical (3D) or polar coordinates (2D) and then inserting the Green's functions.

The correction filter F should tranform the 3D data into 2D data, therefore it can be applied on the Green's function, so that $G^{2D} = F(r,k)G^{3D}$. Thus we have

$$F(r,k) = \frac{G^{2D}}{G^{3D}} = \sqrt{\frac{2\pi r}{k}} e^{i\pi/4} \tag{4.24}$$

The factor $e^{i\pi/4}$ effects a phase shift of $\frac{\pi}{4}$, and the squared term effects a correction of the amplitudes. With $k = \frac{w}{c}$, this can be written as

$$F(r,k) = \sqrt{2rc}\sqrt{\frac{\pi}{w}}\ e^{i\pi/4} = \sqrt{2rc}\ \mathrm{FT}\{\sqrt{t^{-1}}\}, \tag{4.25}$$

where $\mathrm{FT}\{\sqrt{t^{-1}}\}$ is the Fourier transform of the function $\sqrt{t^{-1}}$. We define this now as the phase correction. It is independent of r and easy to implement.

The amplitude correction $\sqrt{2rc}$ depends on the travel distance r. The relation $r = ||x - x_s||$ is only valid for a homogeneous medium, where no reflections can occur. This leads to different practical implementations of the amplitude correction depending on the travel path of the waves. We will here take a look at two different cases.

- **Case 1:** Reflection seismics

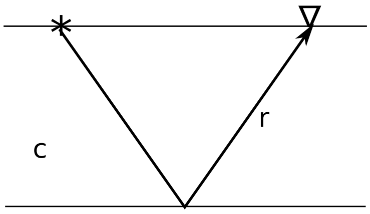


Figure 4.2.: Travel path of a reflected wave

If reflected waves are recorded, their travel path r is in the beginning unknown and can only be calculated if we know the model velocity. With an average velocity c and the recorded time t, the travel path is $r = ct$. We substitute r in the amplitude correction term to eliminate it, so that

$$F_{amp} = \sqrt{2rc} = c\sqrt{2t} \tag{4.26}$$

The amplitude correction factor thus is proportional to the square root of the travel time, which makes the correction quite simple. It works only well in smooth models and for not too complex wave paths.

- **Case 2:** Shallow surface seismics

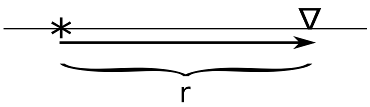


Figure 4.3.: Travel path of a direct wave

For shallow seismic fields, the waves are travelling nearly on the direct wave, so that the travel path r is equal to the offset and therefore known. We can thus eliminate the unknown velocity c in the amplitude correction factor via $c = \frac{r}{t}$:

$$F_{amp} = \sqrt{2rc} = r\sqrt{\frac{2}{t}} \tag{4.27}$$

Even if the spreading correction was initially derived for a homogeneous acoustic medium, this correction for shallow surface seismics works surprisingly well also for elastic surface waves.

5. Full Waveform Inversion of shallow-seismic wavefields

5.1. Motivation

In this chapter, the application of FWI to shallow-seismic surface waves will be described. Shallow seismic refers in this context to a penetration depth of the waves up to 20-30 m. This zone, also called the "critical zone", is the transition zone between earth and atmosphere and therefore highly affected by weathering. This leads to very strong vertical and sometimes lateral variations of visco-elastic material properties that can be in the order of a few 100 %.

The imaging of this "critical zone" is important for different purposes. It is necessary for a geotechnical site characterization, which gives conclusions about the stability of buildings. It is also relevant for hazard analysis. Cavities in the shallow underground can be detected and so a collapse of them can be predicted or even avoided. Additionally, the shear wave velocity in the uppermost 30 m is a measure for the local site amplification due to surface waves. The site amplification indicates how high the amplitudes of surface waves would be at a certain location. If it is known, it can be predicted for an earthquake in which regions the highest damage will occur. Other application fields of shallow seismics can be found among others in hydrology or archaeological prospection.

Shallow seismic wavefields are composed of surface waves, refracted waves and reflected waves (see figure 5.1a)). The refracted waves are the fastest and therefore the first ones that are recorded at the receivers. Reflected waves only occur if there are strong contrasts in the underground. The surface waves are the slowest waves but show the highest amplitudes. They are behaving dispersive (see figure 5.1b)), i.e. higher frequencies are travelling slower than low frequencies.

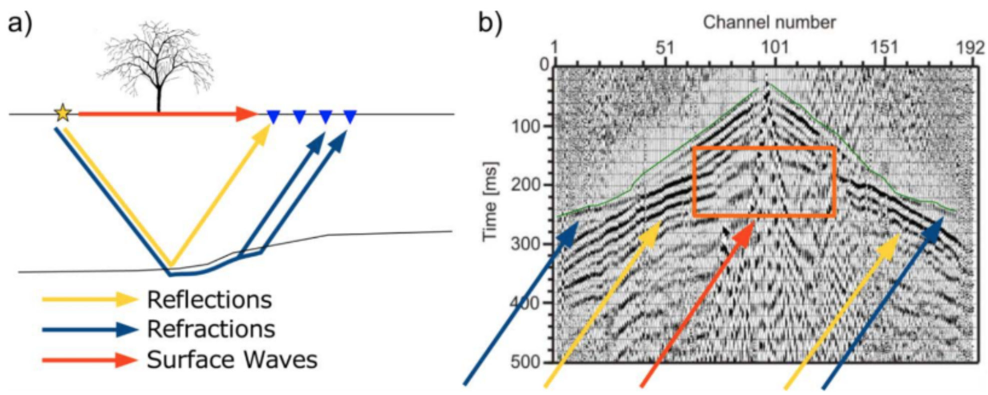


Figure 5.1.: **a)** Different wavetypes occurring in the shallow seismic wavefield. **b)** Seismograms of a shallow seismic wavefield, the arrows are indicating the wavetypes of a).

For geotechnical site characterization, the surface waves are the most suitable because of the following reasons:

- They are easily excited, e.g. by hammer blows or explosions
- They generate the strongest signals in the seismograms
- The penetration depth is up to 30-40 m
- They are highly sensitive to the shear wave velocities

The classical approach to get information about the shallow underground is the inversion of dispersion curves, using only the surface waves. From the dispersion curves, it is possible to get the 1D function $v_s(z)$.
In contrast to this method, a Full Waveform Inversion exploits the full information content of the seismograms, so additionally to the surface waves the refracted and reflected waves are also considered.

5.2. Elastic FWI

For shallow seismic fields, an elastic FWI is necessary. It performs a joint inversion of the velocities v_p, v_s and the density ρ. As misfit function E, the L2-misfit of the normalized seismograms is used:

$$E = \frac{\sum_{i=1}^{N_s} \sum_{j=1}^{N_r} |\hat{s}_{i,j} - \hat{d}_{i,j}|^2}{N_s N_r} \tag{5.1}$$

with the normalized synthetic seismograms $\hat{s}_{i,j} = s_{i,j}/|s_{i,j}|$ and the normalized observed data $\hat{d}_{i,j} = d_{i,j}/|d_{i,j}|$. N_s and N_r are the numbers of sources and receivers, respectively.
For the iterative model updates, the conjugate gradient method is used, where the misfit gradients are calculated with the adjoint state method (see chapter 2). The forward modelling is done with a viscoelastic Finite Difference algorithm.
In the following sections, two challenges are described that have to be handled in order to get results of high quality: The generation of a suitable starting model and the implementation of attenuation in the FWI.

5.2.1. Generation of starting model

The starting model for the inversion has a high influence on the accuracy and convergence of the method. Depending on the starting model, different local minima of the misfit function can be reached instead of the global minimum, that would represent the best solution. We will show this effect for a specific example, as well as possibilities to improve the starting model.

As example, we consider a homogeneous halfspace with the velocity $v_{s,1} = 200\,\frac{\mathrm{m}}{\mathrm{s}}$ and a thin layer in a depth between 3 m and 5 m with a velocity of $v_{s,2} = 400\,\frac{\mathrm{m}}{\mathrm{s}}$ (see table 5.1). Figure 5.2 shows the misfit function for this model. As misfit, the L2-norm of the residuals of true and synthetic data is used.

Table 5.1.: S-wave velocity model: The depth z indicates the upper edge of the layers.

z in m	v_s in m/s
0	200
3	400
5	200

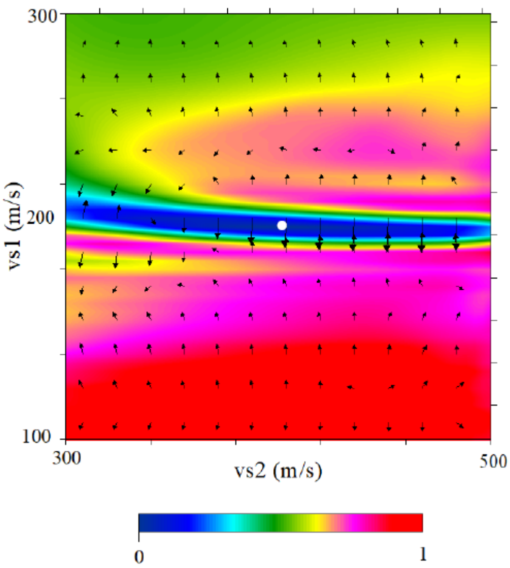


Figure 5.2.: Misfit function in dependence of the two S-wave velocities $v_{s,1}$ and $v_{s,2}$. The interfaces are at the same depth as in the model (table 5.1). The white point indicates the true model parameters, the arrows are showing the directions of the gradient.

In figure 5.2, we can see that there is not only one minimum for the true model parameters, but a wide one, that is almost independent of the velocity $v_{s,2}$. The arrows show, in which direction the parameters would be changed by the steepest gradient method. We can see that if for example $v_{s,1}$ is chosen slightly too high in the starting model, the parameters would change in the wrong direction, so that the final model would be further away from the true one than the starting model.

To find an appropriate starting model, there exist different possibilities. One of them is to apply a frequency filter on the data. Figure 5.3 shows the misfit function after a lowpass filter with 15 Hz has been applied on the data. The minimum is now much wider, so that the starting model can differ more from the true one without resulting in a wrong local minimum.

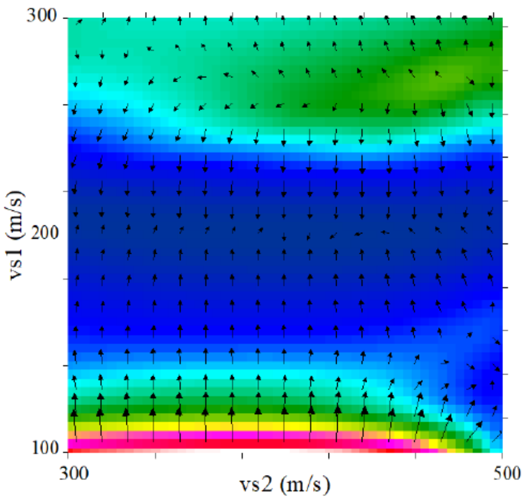


Figure 5.3.: Misfit function for filtered data, as filter a 15 Hz lowpass was used

The inversion result can therefore be improved by performing first an inversion for the filtered data, and then using the result as starting model for the data without filtering. It is also possible to use different filters, starting with a small bandwidth and augmenting it slowly until the whole frequency content of the data is used.

A disadvantage of this method is that it is very time consuming. Therefore, other methods to derive a starting model might be more effective. A second method is to build the starting model via an inversion of the dispersion curves. This can be done in one dimension, resulting in a velocity function $v_s(z)$. This result can then be used as starting model for the FWI. Figure 5.4 shows the corresponding misfit function. We can see that the wide global minimum is now reached for all starting models in the showed range.

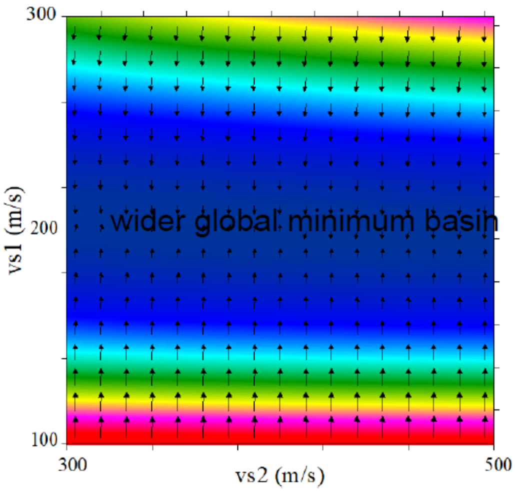


Figure 5.4.: Misfit function with inversion of dispersion curves

??? was genau zeigt Abb.?

5.2.2. Attenuation

A second challenge for FWI is the attenuation of seismic data. In a viscoelastic medium, we have the effect of intrinsic attenuation. Wave energy is transformed into heat or into the motion of particles, which results in a loss of amplitude depending on the travel distance of the wave. Additionally, the attenuation is frequency-dependent, the amplitudes of high frequencies are damped more than those of lower frequencies. This leads to a dispersion effect visible in the seismograms.

To compensate the decrease of amplitudes with the offset, normalized seismograms are used to calculate the misfit function (see equation 5.1). To model the frequency loss with the distance correctly, it is necessary to do a source signal inversion, where the damping parameter Q is considered. This damping parameter has to be estimated before on the basis of a-priori information. When Q is known, it can be used for a viscoelastic forward modelling, which is able to model the dispersion as a consequence of attenuation correctly. We thus perform an elastic FWI, where only the forward modelling is calculated for a viscoelastic medium. Until now, it is not possible to invert also for the damping parameters Q.

A. Appendix

A.1. The Green's function

In the following, we consider the scalar acoustic wave equation of the form

$$\left[\frac{\partial^2}{\partial t^2} - c^2(x)\Delta\right] u(x,t) = f(x,t) \tag{A.1}$$

with $u(x,0) = \frac{\partial u}{\partial t}(x,0) = 0$. Here, u is the pressure, c the sound velocity and f the source function. In general, we have three space dimensions, so $x \in \mathbb{R}^3$ and $\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}$.

The Green's function $G(x,y,t)$ of this differential equation is its impulse response, which means it is the solution of the equation with a delta-function as source function:

$$\left[\frac{\partial^2}{\partial t^2} - c^2(x)\Delta_x\right] G(x,y,t) = \delta(x-y)\delta t \tag{A.2}$$

The source function describes a delta-impulse at the location $x = y$ and at the time $t = 0$. Therefore, the Green's function depends on the receiver position x, the source position y and the time t.

If we know the Green's function, we can calculate the pressure field for any arbitrary source function by convolving the Green's function with the source function and integrating over the whole model space:

$$u(x,t) = \int_0^t \int_{\mathbb{R}^3} G(x,y,t-s) f(y,s)\; dy\; ds \tag{A.3}$$