The adjoint method in seismology – I. Theory

A. Fichtner * , H.-P. Bunge, H. Igel,

Department of Earth and Environmental Sciences, Ludwig-Maximilians University Munich, Theresienstrasse 41, D-80333 Munich, Germany

Abstract

The adjoint method is a powerful tool in geophysics that permits the computation of the exact first derivative of a physical observable or an associated objective function with respect to its parameters. Typical observables are displacement fields and flow patterns at the surface. Possible geophysical parameters are density, viscosity or elastic coefficients. When the observable can be modelled by solving a differential equation, the computation of the derivative only consists in solving the forward problem and its adjoint problem. Therefore, the adjoint method is far more efficient than any finite difference approximation. Here we present a mathematical formalism that generalises the derivation of the adjoint problem. In order to connect to work by Tarantola (1984, Geophysics, 49(8), 1259) we first give a derivation of the adjoint equations for the scalar wave equation in two dimensions. As objective function we choose the time integral over the quadratic difference between the modelled wave field and real data. In this case the adjoint problem coincides with the original forward problem, the only difference being that the adjoint field satisfies terminal rather than initial conditions. A numerical example in two dimensions demonstrates that the adjoint field focusses near the location of a parameter perturbation at the same time when the original wavefront reaches that location. Based on this simple example, we introduce a generalised formalism for the adjoint method. It is independent of the existence of Green's functions and their spatio-temporal reciprocity relations. Moreover, the formalism applies to non-linear equations such as the Navier-Stokes equations. This may become important in mantle flow reconstructions. The source term of the adjoint equations depends only on the specific objective function. Choosing the objective function to coincide with the observable itself, allows us the computation of Jacobians, i.e., the derivative of the observable with respect to the model parameters. To demonstrate the consistency of our formalism with earlier analyses, we consider the anisotropic elastic wave equation with attenuation, which is of major interest in seismology.

Key words: Inversion, Fréchet derivative, waveform analysis, sensitivity
PACS: 91.35.Pn, 91.30.Fn
1991 MSC: 47A05, 49N45

1 Introduction

Determination of the structure and dynamics of the Earth's deep interior is one of the principal objectives of geophysics. Efforts of imaging the Earth on a global scale, early on lead to radial density and velocity models deduced primarily from global observations of the arrival times of seismic phases (Dziewonski et al., 1975; Dziewonski & Anderson, 1981; Kennett & Engdahl, 1991; Kennett et al., 1995). These models are consistent with petrologic mantle models and form the basis of three-dimensional global tomographic images of the Earth's interior. Following the pioneering work by Dziewonski (1977, 1984), Masters et al. (1996) inverted for a shear-velocity model in the mantle combining surface wave, free oscillation and traveltime data. Similar models were obtained by Grand (1994) and Grand et al. (1997). Mégnin and Romanowicz (2000) employed an asymptotic coupling theory for normal modes in order to invert body, surface and higher-mode waveforms for shear heterogeneity in the mantle. Body and Rayleigh waves as well as normal mode splitting functions have been used by Ritsema and van Heijst (2000). Tomographic images of the lowermost mantle have been obtained by Kárason and van der Hilst (2001) by including differential traveltimes of core phases. Gorbatov and Kennett (2003) and Kennett and Gorbatov (2004) jointly inverted the arrival-times of P and S waves for bulk-sound and shear wavespeed anomalies in the mantle, thus increasing the interpretability in terms of temperature and compositional variations. These studies have greatly improved our understanding of three-dimensional heterogeneity; and the large-scale mantle heterogeneity structure is now well agreed upon (Becker & Boschi, 2002).
In a related effort geodynamicists have demonstrated a strong correlation between the history of subduction and large-scale seismic mantle heterogeneity structure (Richards & Engebretson, 1992; Lithgow-Bertelloni & Richards, 1998). This latter insight has led to the construction of so-called mantle circulation models (Bunge et al., 1998, 2002). The dynamic processes of mantle convection manifest themselves as lateral variations in density, temperature and composition, which map into the visco-elastic structure of the mantle. This means that temperature and density variations correspond to variations in seismic velocities (Brown & Shankland, 1981; Duffy & Ahrens, 1992; Mattern et al., 2005). Therefore, the analysis of seismic waves allows us to infer flow patterns in the mantle (van der Hilst et al., 1997; Kárason & van der Hilst, 2000)
It is indisputable, that the existing tomographic images of the mantle successfully contributed to the understanding of the planet's dynamics. Still, the inversions are based on substantially simplified forward models, namely ray theory and finite normal mode summations. Ray theory is only applicable

* Corresponding author.
Email address: andreas.fichtner@geophysik.uni-muenchen.de (A. Fichtner).

to the arrival times of high frequency waves, therefore significantly reducing the amount of exploitable information. Normal mode approximations rely on smoothly varying structure and long period waveforms, resulting in a limitation of resolution.

The fact that today's computational power is sufficient to accurately solve the wave equation in realistic Earth models (e.g. Igel et al., 1995; Komatitsch et al., 2000) suggests that the next step in seismic inversion consists in replacing the approximate forward models by the exact forward model. This may allow us to invert for seismic waveforms with shorter periods. Intuitively, one expects that the resulting increase of exploited information translates to an increase of resolution especially in poorly sampled regions.

There exist different conceptions of what it means to solve an inverse problem. In probabilistic inverse theory (Tarantola, 1987) the solution of the inverse problem is defined as a marginal probability density in the model space. Unless the probability density is very simple, it can only be characterized by exploring the model space, usually on the basis of Monte Carlo methods (e.g. Press, 1968). Though very general and elegant, this approach suffers from the large number of forward problem evaluations necessary to perform the model space exploration. Therefore the process of solving the inverse problem is in practice often equated with the minimisation of the difference between observed and synthetic data with respect to the model parameters. In the context of waveform inversion, the solution would thus be defined as the model $\mathbf{p}_{min}$ that minimizes the difference between an observed waveform $\mathbf{u}_0$ and a synthetic waveform $\mathbf{u}$. This difference can be quantified through an objective function $\mathfrak{E}$, which may additionally depend on the model parameters $\mathbf{p}$ and a priori parameters $\mathbf{p}_0$.

In this context the total derivative of $\mathfrak{E}$ with respect to $\mathbf{p}$, denoted by $D_p\mathfrak{E}$, becomes important. It may be used for an inversion based on a gradient method and for sensitivity and resolution analyses. The major complication is that the computation of $D_p\mathfrak{E}$ requires the computation of $D_p\mathbf{u}$, i.e., the total derivative of the wavefield $\mathbf{u}$ with respect to the model parameters. Due to the very large size of the model space it is practically infeasible to obtain this quantity by classical finite differencing techniques.

An elegant and physically insightful solution to this problem is the *adjoint method*. It allows us to compute the derivative with respect to the parameters by combining the synthetic forward wavefield and an adjoint wavefield governed by a set of adjoint equations and adjoint subsidiary conditions. This concept was introduced by Tarantola (1984, 1988) into the field of seismology. It forms the basis of numerical studies by Gauthier et. al (1986) and applications to the inversion for 1-D and 2-D structure from marine reflection seismograms (Crase et al., 1990; Igel et al., 1996). Recently, the adjoint method was used in the context of finite-frequency traveltime kernels (Tromp et al., 2005). One of the principal characteristics of the adjoint problem in seismology is time reversal, meaning that the adjoint problem consists of a propagation of the observed waveform residuals backward in time and from the receiver to

the source. Time reversals are relatively common in geophysics and in particular in the field of reverse-time migration (e.g. Baysal, 1983). This technique however, focusses only on the imaging of structures that, for a given seismic signal frequency, appear as a discontinuity. Moreover, reverse-time migration suffers from two significant deficiencies: it applies the acoustic approximation and unphysical variants of the wave equation; thus, for example, it neglects radiation patterns of both the field emitted by the source and the secondary field set off by a perturbation in density or elastic parameters. Adjoint problems have already been used in other branches of the earth sciences such as meteorology (Talagrand & Courtier, 1987) and geodynamics (Bunge et al., 2003). Even though the principal ideas are identical, the differences between the physics of the Navier-Stokes equations and the wave equation renders a one-to-one translation of methods from geodynamics to seismology impossible.

In the present paper we derive the adjoint problem for the case of a simple two-dimensional scalar wave equation and a least squares objective function. Based on this example we introduce a formalism that allows to generalise the adjoint method to arbitrary differential operators and objective functions. The derivatives of the objective functions are exact. Moreover, it becomes clear that it is possible to extend the method to non-linear equations and to compute wave field Jacobians by simply choosing specific objective functions. As an application relevant for seismology, we consider the anisotropic wave equation with attenuation. The resulting adjoint problem coincides with the one found by Tarantola (1988), which demonstrates the consistency of our approach.

The results presented in this paper form the basis of a second paper on the adjoint method in seismology. The theory introduced in the following sections will there be used in order to deduce expressions for waveform sensitivity kernels and information about the physical meaning of the first derivative. A waveform inversion procedure applicable on continental or global scales is the long-term objective to which this study hopes to contribute.

2 Preliminaries

The adjoint method can be described most elegantly and most efficiently with a modern mathematical notation which we will briefly introduce here (e.g. Kantorowitsch & Akilow (1964)). Consider an operator P mapping an element x of a space X to an element y of another space Y, i.e., $P(x) = y$. The operator P may for example be an ordinary scalar function or a tensor of any order. The elements x and y can be scalars, tensors or even other operators. A *linear*

operator U satisfying the relation

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[P(x + \varepsilon x_0) - P(x)] = U(x_0) \tag{1}$$

for all $x_0 \in X$, is called the *first derivative* of P with respect to x in the direction x_0. Symbolically, we write $U(x_0) = DP(x)(x_0)$. This derivative is called the Gâteaux derivative or weak derivative. If it exists uniformly with respect to the differentiation directions x_0, it is termed the Fréchet derivative or strong derivative. A justification for this notation, which is in contrast to the commonly used dy/dx, is that a quotient of two elements $y \in Y$ and $x \in X$ may not be defined. This is the case for example if x is a vector in $\mathbb{R}^n$ or a distributed variable that vanishes at some point. Moreover, the generalised derivative has always a direction x_0 that has to be part of the notation. When f is a function depending on a vector $\mathbf{x} \in \mathbb{R}^3$, we may alternatively write $Df(\mathbf{x})(\mathbf{x}_0) = (\mathbf{x}_0 \cdot \nabla)f(\mathbf{x})$. One may extend the well-known chain rule to the case of composed operators. Assuming that for an operator R the expression $R(x)$ is given by $R(x) = P(Q(x))$, i.e., the composition of the operators P and Q, the generalised chain rule states that the derivative of R with respect to x in the direction x_0 is given by

$$DR(x)(x_0) = DP(Q(x))(y_0) \quad \text{with} \quad y_0 = DQ(x)(x_0). \tag{2}$$

When P takes more than one variable, e.g. $y = P(x_1, x_2)$, with x_1 and x_2 elements of the spaces X_1 and X_2, respectively, we define

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[P(x_1 + \varepsilon x_1', x_2) - P(x_1, x_2)] =: U_1(x_1') =: \partial_{x_1} P(x_1, x_2)(x_1')\,, \tag{3}$$

where $=:$ denotes equality by definition. If U_1 exists and if it is linear with respect to x_1', it is called the first partial derivative of P with respect to x_1 in the direction of x_1'. The partial derivative $\partial_{x_2} P(x_1, x_2)(x_2')$ is defined in analogy to equation (3).
When the function u is a physical observable (e.g. a seismic wavefield) depending on a parameter set p (e.g. density and elastic moduli), then Taylor's theorem yields a physical interpretation of the direction of differentiation: $u(p + q) - u(p) \doteq Du(p)(q)$. Hence, the first derivative with respect to the parameters p in the direction q is, correct to first order in $||q||$, the difference between $u(p)$ and $u(p + q)$.

3 The adjoint method for the two-dimensional scalar wave equation and the least squares objective function

This section gives a simple example that demonstrates the principal idea of the adjoint method. It proceeds in two steps. First, the definition of an objective

function $\mathfrak{E}$ that depends on the physical observable u. Second, the derivation of the adjoint problem of the two-dimensional scalar wave equation and its subsidiary conditions. The total derivative of $\mathfrak{E}$ with respect to the parameter set $\mathbf{p}$, can then be expressed in terms of u and the solution of the adjoint problem ψ. In our case, u is a scalar wavefield and $\mathbf{p}$ comprises the density and a shear modulus distributions, i.e., $\mathbf{p}(\mathbf{x}) = (\rho(\mathbf{x}), \mu(\mathbf{x}))$.
An objective function is a functional that acts on a physical observable. More specifically, it can be used to quantify the difference between synthetic data u and observed data u_0. Due to its simplicity the least squares objective function plays a central role in many physical applications. Let $u_0(\boldsymbol{\xi}, t)$ be scalar data observed at the point $\mathbf{x} = \boldsymbol{\xi} \in G \subset \mathbb{R}^2$ and in the time interval $t \in [t_0, t_1]$. G is the region in which u is defined. Then the least squares objective function is given by

$$\mathfrak{E}(u) = \frac{1}{2} \int_{t=t_0}^{t_1} [u(\boldsymbol{\xi}, t) - u_0(\boldsymbol{\xi}, t)]^2 \, dt = \frac{1}{2} \int_{t=t_0}^{t_1} \int_G [u(\mathbf{x}, t) - u_0(\mathbf{x}, t)]^2 \delta(\mathbf{x} - \boldsymbol{\xi}) \, d^2\mathbf{x} \, dt \,, \tag{4}$$

where u is a synthetic wavefield computed from the two-dimensional wave equation with parameters $\mathbf{p} = (\rho, \mu)$. We may write the integral (4) in the condensed form

$$\mathfrak{E}(u) = \langle 1, f(u) \rangle \quad \text{with} \quad f(u) := \frac{1}{2}[u(\boldsymbol{\xi}, t) - u_0(\boldsymbol{\xi}, t)]^2 \delta(\mathbf{x} - \boldsymbol{\xi}) \,, \tag{5}$$

which will be useful in section (4). We are interested in the total derivative of $\mathfrak{E}$ with respect to the parameters $\mathbf{p}$ in the direction of $\mathbf{q} = (\rho', \mu')$. The application of the chain rule yields

$$D_p \mathfrak{E}(u)(\mathbf{q}) = \partial_u \mathfrak{E}(u)(D_p u(\mathbf{q})) = \langle 1, \partial_u f(u)(D_p u(\mathbf{q})) \rangle \,. \tag{6}$$

The obvious problem is the presence of $D_p u(\mathbf{q})$ in equation (6). Due to the size of the parameter space (all possible distributions of μ and ρ), this quantity can usually not be approximated through finite differencing. The objective of the adjoint method is the elimination of $D_p u(\mathbf{q})$ from equation (6).
Now consider the two-dimensional scalar wave equation in the domain G with boundary ∂G, given by

$$\rho(\mathbf{x}) \partial_t^2 u(\mathbf{x}, t) - \nabla \cdot (\mu(\mathbf{x}) \, \nabla u(\mathbf{x}, t)) = g(\mathbf{x}, t) \tag{7a}$$

for $\mathbf{x} \in G \subset \mathbb{R}^2$, $t \in [t_0, t_1]$ and complemented by the set of subsidiary conditions:

$$u(\mathbf{x}, t)|_{\mathbf{x} \in \partial G_1} = 0 \,, \qquad \text{Cauchy condition} \tag{7b}$$

$$\mathbf{n} \cdot \nabla u(\mathbf{x}, t)|_{\mathbf{x} \in \partial G_2} = 0 \,, \qquad \text{Neumann condition} \tag{7c}$$

$$u(\mathbf{x}, t)|_{t=t_0} = \partial_t u(\mathbf{x}, t)|_{t=t_0} = 0 \,, \qquad \text{initial conditions} \tag{7d}$$

with $\partial G_1 \cup \partial G_2 = \partial G$. All vector quantities in these equations, including the gradient operator, are two-dimensional. Noting that $D_p = D_\mu + D_\rho$, we can differentiate equation (7a) with respect to the parameters $\mathbf{p}$,

$$\rho' \partial_t^2 u - \nabla \cdot (\mu' \nabla u) + \rho \partial_t^2 v - \nabla \cdot (\mu \nabla v) = 0\,, \tag{8}$$

with $v := D_p u(\mathbf{q})$. Equation (8) is homogeneous because the source term g is independent of the parameters $\mathbf{p} = (\rho, \mu)$. We now introduce an arbitrary, but sufficiently nice *test function* ψ. Combining ψ and equation (8) by means of the integral $\langle .\,, .\rangle$ yields

$$\langle \psi, \rho' \partial_t^2 u - \nabla \cdot (\mu' \nabla u)\rangle + \langle \psi, \rho \partial_t^2 v - \nabla \cdot (\mu \nabla v)\rangle = 0\,. \tag{9}$$

The next task is to transform the second summand in equation (9) such that ψ and v reverse their positions. In doing so we will have to subject the test function ψ to various conditions. Finally, a complete set of conditions will lead to a precise determination of ψ, therefore transforming it from the originally arbitrary test function ψ into the well-defined *adjoint wavefield*. We start with the term involving $\partial_t^2 v$. It can easily be transformed using a double integration by parts.

$$\langle \psi, \rho \partial_t^2 v\rangle = \int_G \rho \psi \partial_t v \, d^2\mathbf{x} \,|_{t=t_0}^{t_1} - \int_G \rho v \partial_t \psi \, d^2\mathbf{x} \,|_{t=t_0}^{t_1} + \langle v, \rho \partial_t^2 \psi\rangle. \tag{10}$$

The homogeneous initial conditions for $v = D_p u(\mathbf{q})$, which follow from the differentiation of the initial conditions (7d) for u, imply that we can eliminate the first two terms on the right-hand side of (10) by imposing the *terminal conditions* $\psi(\mathbf{x}, t)|_{t=t_1} = 0$ and $\partial_t \psi(\mathbf{x}, t)|_{t=t_1} = 0$ upon ψ. Then we obtain

$$\langle \psi, \rho \partial_t^2 v\rangle = \langle v, \rho \partial_t^2 \psi\rangle. \tag{11}$$

Similarly, the term $\langle \psi, \nabla\cdot(\mu\nabla v)\rangle$ can be transformed with the two-dimensional version of Gauss' theorem and the differentiated boundary conditions (7b) and (7c). Using the identity

$$\nabla \cdot (\mu\psi\nabla v) - \nabla \cdot (\mu v \nabla \psi) = \psi \nabla \cdot (\mu \nabla v) - v \nabla \cdot (\mu \nabla \psi)\,, \tag{12}$$

we find

$$\langle \psi, \nabla\cdot(\mu\nabla v)\rangle = \int_{t=t_0}^{t_1} \int_{\partial G_1} (\mu\psi\nabla v)\cdot\mathbf{n}\, ds\, dt - \int_{t=t_0}^{t_1} \int_{\partial G_2} (\mu v \nabla\psi)\cdot\mathbf{n}\, ds\, dt + \langle v, \nabla\cdot(\mu\nabla\psi)\rangle\,, \tag{13}$$

where ds is a line element and $\mathbf{n}$ is the outward-pointing normal on the curves ∂G_1 and ∂G_2, respectively. By imposing the two additional conditions $\psi(\mathbf{x}, t)|_{\mathbf{x}\in\partial G_1} = 0$ and $\mathbf{n}\cdot\nabla\psi(\mathbf{x}, t)|_{\mathbf{x}\in\partial G_2} = 0$ upon ψ, equation (13) reduces to

$$\langle \psi, \nabla \cdot (\mu\nabla v)\rangle = \langle v, \nabla \cdot (\mu\nabla\psi)\rangle\,. \tag{14}$$

Without imposing any additional constraints on ψ, similar transformations lead to

$$\langle \psi, \rho' \partial_t^2 u - \nabla \cdot (\mu' \nabla u) \rangle = -\langle \rho', \partial_t \psi \, \partial_t u \rangle + \langle \mu', (\nabla u) \cdot (\nabla \psi) \rangle \, . \tag{15}$$

We may now rewrite equation (9) as

$$\langle v, \rho \partial_t^2 \psi - \nabla(\mu \nabla \psi) \rangle - \langle \rho', \partial_t \psi \, \partial_t u \rangle + \langle \mu', (\nabla u) \cdot (\nabla \psi) \rangle = 0 \, . \tag{16}$$

Remembering that $v = D_p u(\mathbf{q})$ is the derivative that we wish to eliminate from the derivative of $\mathfrak{E}$ (see equation (6)), we add the homogeneous equation (16) to equation (6),

$$D_p \mathfrak{E}(u)(\mathbf{q}) = \langle v, \partial_u f + \rho \partial_t^2 \psi - \nabla(\mu \nabla \psi) \rangle - \langle \rho', \partial_t \psi \, \partial_t u \rangle + \langle \mu', (\nabla u) \cdot (\nabla \psi) \rangle \, . \tag{17}$$

It is possible to eliminate v by imposing one last condition upon ψ, namely

$$\rho(\mathbf{x}) \partial_t^2 \psi(\mathbf{x}, t) - \nabla \cdot (\mu(\mathbf{x}) \, \psi(\mathbf{x}, t)) = -\partial_u f \tag{18a}$$

At this point we can state our final result: Given that the function ψ satisfies equation (18a) for $\mathbf{x} \in G \subset \mathbb{R}^2$, $t \in [t_0, t_1]$ and the set of subsidiary conditions:

$$\psi(\mathbf{x}, t)|_{\mathbf{x} \in \partial G_1} = 0 \, , \qquad \text{Cauchy condition} \tag{18b}$$

$$\mathbf{n} \cdot \nabla \psi(\mathbf{x}, t)|_{\mathbf{x} \in \partial G_2} = 0 \, , \qquad \text{Neumann condition} \tag{18c}$$

$$\psi(\mathbf{x}, t)|_{t=t_1} = \partial_t \psi(\mathbf{x}, t)|_{t=t_1} = 0 \, , \qquad \text{terminal conditions} \tag{18d}$$

then the total derivative of the objective function $\mathfrak{E}$ with respect to the model parameters $\mathbf{p} = (\rho, \mu)$ in the direction of $\mathbf{q} = (\rho', \mu')$ is given by

$$D_p \mathfrak{E}(u)(\mathbf{q}) = -\langle \rho', \partial_t \psi \, \partial_t u \rangle + \langle \mu', (\nabla u) \cdot (\nabla \psi) \rangle \, . \tag{19}$$

The set of equations (18a) to (18d) is referred to as the *adjoint problem* of equations (7a) to (7d). Combining u and the adjoint field ψ according to equation (19) gives the exact derivative of the objective function $\mathfrak{E}$. What makes this method attractive is the fact that the adjoint problem is very similar to the original problem, meaning that it also consists in solving a two-dimensional wave equation subject to a set of subsidiary conditions. The spatial boundary conditions translate one-to-one, whereas the temporal boundary conditions translate from an initial condition in the original problem to a terminal condition in the adjoint problem. This unusual terminal condition can be interpreted by considering the source term of the adjoint equations, namely

$$-\partial_u f = -[u(\mathbf{x}, t) - u_0(\mathbf{x}, t)] \, \delta(\mathbf{x} - \boldsymbol{\xi}) \, . \tag{20}$$

The source acts at a single point and its time function consists in the negative linear residuals, i.e., the linear difference between the synthetic and observed data. Therefore, solving the adjoint problem can be interpreted as a propagation of the residuals backward in time, meaning from $t = t_1$ to $t = t_0$. As will

become clear in the next section, the residuals will focus in the region where they came from: a parameter perturbation of the assumed model parameters $\mathbf{p}$ with respect to the true model parameters $\mathbf{p}_0$. Note that the source of the forward wavefield is completely absent in the adjoint problem. It merely enters implicitly via u.

The two-dimensional scalar wave equation is well suited for an illustration of the physical meaning of wave equation adjoints in general because the principal concept translates to more complex cases which include elasticity and anisotropy.
The numerical examples in this section are based on a finite differences solution of equations (7a) to (7d). A free surface condition is implemented at the top ($z = 0\,\mathrm{km}$), whereas the bottom boundary ($z = 150\,\mathrm{km}$) is rigid. Periodic boundary conditions are implemented on the left and right boundaries. The parameter model $\mathbf{p}_0 = (\rho_0, \mu_0)$ used to generate the data u_0 is homogeneous in both density ($\rho = 3.0 \cdot 10^3\ \mathrm{kg\,m^{-3}}$) and shear modulus ($\mu = 75 \cdot 10^9\ \mathrm{N\,m^{-2}}$) with the exception of one single cell located at $(x, z) = (150, 70)\,\mathrm{km}$ where the density is increased to a value of $\rho = 3.5 \cdot 10^3\,\mathrm{kg\,m^{-3}}$. Therefore, the shear velocity equals $v_s = 5\,\mathrm{km\,s^{-1}}$ almost everywhere in the parameter model. The reference model $\mathbf{p}$ is completely homogeneous. Hence, the observed residuals arise merely from a single-cell perturbation of one model with respect to the other. The two wavefields $u(\mathbf{p})$ and $u(\mathbf{p}_0)$ are both recorded at 150 evenly spaced receivers on the free surface ($z = 0$). This relatively high number of receivers will mostly be unrealistic in practical applications but it can well be justified for the purpose of illustration. Note that due to linearity an objective function consisting of a sum of time-integrated squared residuals,

$$\mathfrak{E}(u) = \frac{1}{2} \sum_{i=1}^{N} \int_{t=t_0}^{t_1} [u(\mathbf{p}; \boldsymbol{\xi_i}, t) - u(\mathbf{p}_0; \boldsymbol{\xi_i}, t)]^2 \, dt \,, \tag{21}$$

where $\boldsymbol{\xi}_i$ are the locations of the N receivers, simply translates to a superposition of adjoint sources and therefore to a superposition of adjoint wavefields. Since the source of the forward wavefields does not explicitly enter the adjoint equations, there are no restrictions whatsoever on that source. It may be a spatially extended source, a point source or a number of point sources. Here, for simplicity, we used a single point source located at $(x, z) = (150, 1)\,\mathrm{km}$, i.e., close to the free surface and directly above the density perturbation. The source radiates a Ricker wavelet with a dominant frequency of 0.3 Hz.

Figure (1) shows snapshots of the forward field $u(\mathbf{p}; \mathbf{x}, t)$ (left) and the adjoint field $\psi(\mathbf{x}, t)$ (right) at times $t = (100, 185, 300, 400)\,\mathrm{s}$. The colour scales are not uniform but individually adjusted in order to emphasise the geometries of the wavefields at different times. Before $t = 150\,\mathrm{s}$ the two wavefields overlap only weakly, therefore resulting in a small contribution to the derivative of the objective function derivative, which is $D_p \mathfrak{E}(u)(\rho) = -\langle \rho', \partial_t \psi \, \partial_t u \rangle$. Around $t = 185\,\mathrm{s}$ the adjoint wavefield focusses near the location of the density per-

Fig. 1. Snapshots of the forward wavefield $u(\mathbf{p};\mathbf{x},t)$ (left) and the adjoint wavefield $\psi(\mathbf{x},t)$ (right). The major contribution to the derivative $D_p\mathfrak{E}(u)(\mathbf{q})$ arises from a focussing of the adjoint field near the location of the density perturbation at $(x,z) = (150,70)$ km (indicated by '$\times$') and a simultaneous passage of the wavefront of the forward field through that point. The two fields overlap only weakly before and after the focussing, therefore leading to significantly smaller contributions. A single point source indicated by 'o' is at $(x,z) = (150,1)$ km.

Fig. 2. Total derivative of the least squares objective function $\mathfrak{E}$ with respect to density. A clearly visible peak is located near the density perturbation at $(x,z) = (150,70)$ km. Additional non-zero contributions extend to the surface in the form of narrow branches.

turbation, namely at $(x,z) = (150,70)$ km. At the same time the wavefront of u passes through this point. Consequently, the product $\partial_t u\,\partial_t\psi$ becomes large, resulting in a significant contribution to $D_p\mathfrak{E}(u)(\mathbf{q})$. Subsequently, the adjoint field de-focusses and finally disappears, as required by the terminal conditions.

The resulting total derivative of $\mathfrak{E}$ with respect to ρ is shown in figure 2. Its mayor contribution is located near the density perturbation, as expected. However, various additional branches extend up to the surface. Moreover, the derivative peak is not restricted to the one grid cell where the density perturbation is situated. This effect is related to the finite width of the signal.
It is straightforward to repeat the above example with a shear modulus perturbation rather than a density perturbation. Even though the equation for the derivative with respect to μ is different, the fundamental effects remain unchanged and the patterns of the derivative differ only in details.
The derivative $D_\rho\mathfrak{E}(u)(\rho')$ provides information on the first-order changes in density that we have to apply to our density model in order to obtain the perturbed model, which in practice is the Earth model that we wish to invert for. However, a certain waveform residual caused by a pure density perturbation may also by explained by a shear modulus perturbation. Mathematically this phenomenon manifests itself by a mapping of a density perturbation into the shear modulus derivative and vice versa.

4 A generalisation of the adjoint method

For various reasons it is desirable to obtain a generalisation of the adjoint method, outlined in the previous section for the specific case of the two-dimensional scalar wave equation and the least squares objective function.

First of all, such a generalisation will provide deeper insight into the mathematical structure of the problem, therefore potentially leading to new applications. Also, a condensed notation will allow us to treat more complex problems.

4.1 Generalised notation

First, it is necessary to introduce the condensed notation already mentioned: It proves to be efficient to replace an explicit differential equation by an abstract operator $\mathbf{L}$. It maps a physical observable $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ to a right-hand side $\mathbf{g}(\mathbf{x}, t)$. The operator $\mathbf{L}$ may itself depend on a set of parameters $\mathbf{p}$ as well as on the spatial variable $\mathbf{x}$ and the time variable t. In symbols:

$$\mathbf{L}(\mathbf{u}; \mathbf{p}, \mathbf{x}, t) = \mathbf{g}(\mathbf{x}, t)\,. \tag{22}$$

We use bold face symbols to indicate that they are (potentially) vector quantities. The dependence of $\mathbf{L}$ on the parameters $\mathbf{p}$ implies a dependence of $\mathbf{u}$ on $\mathbf{p}$. Therefore, it is more precise to write $\mathbf{u} = \mathbf{u}(\mathbf{p}; \mathbf{x}, t)$. As an example we may once more consider the two-dimensional scalar wave equation. Its corresponding operator is

$$L([\,.\,]; \mathbf{p}, \mathbf{x}, t) = \rho \partial_t^2 [\,.\,] - \nabla \cdot (\mu \nabla [\,.\,])\,, \tag{23}$$

where $[\,.\,]$ indicates the position of the function to which the operator can be applied. The parameters $\mathbf{p}$ are μ and ρ. Using this notation, we can rewrite the two-dimensional scalar wave equation in the form $L(u; \mathbf{p}, \mathbf{x}, t) = g(\mathbf{x}, t)$.
A special notation has already been introduced for the least squares objective function. Now, we will consider an arbitrary objective function that in addition to $\mathbf{u}$ also depends on the model parameters $\mathbf{p}$, i.e., $\mathfrak{E} = \mathfrak{E}(\mathbf{u}, \mathbf{p})$. Moreover, we assume that $\mathfrak{E}$ can be expressed as

$$\mathfrak{E}(\mathbf{u}, \mathbf{p}) = \langle 1, f(\mathbf{u}, \mathbf{p}) \rangle\,, \tag{24}$$

where f is an adequate scalar function and $\langle\,.\,,\,.\,\rangle$ a bilinear form, i.e., an expression that is independently linear in both arguments. Equations (4) and (5) define one possible bilinear form.

4.2 The adjoint method

We are interested in the total derivative of an objective function $\mathfrak{E}(\mathbf{u}, \mathbf{p}) = \langle 1, f(\mathbf{u}, \mathbf{p}) \rangle$ with respect to the model parameters $\mathbf{p}$ in some direction $\mathbf{q}$. Invoking the chain rule gives

$$D_p \mathfrak{E}(\mathbf{u}, \mathbf{p})(\mathbf{q}) = \langle 1, \partial_u f(\mathbf{u}, \mathbf{p})(\mathbf{v}) \rangle + \langle 1, \partial_p f(\mathbf{u}, \mathbf{p})(\mathbf{q}) \rangle\,, \tag{25}$$

with $\mathbf{v} := D_p\mathbf{u}(\mathbf{p};\mathbf{x},t)(\mathbf{q})$. As already mentioned in the section on the two-dimensional scalar wave equation, the presence of $D_p\mathbf{u}$ in equation (25) is problematic. Due to time-consuming forward problem solvers and very large model spaces it will often be unfeasible to approximate this derivative by means of finite differencing. To eliminate $D_p\mathbf{u}$ from the expression for $D_p\mathfrak{E}$ we first rewrite equation (25) as

$$D_p\mathfrak{E}(\mathbf{u},\mathbf{p})(\mathbf{q}) = \langle \mathbf{v}, \partial_u f(\mathbf{u},\mathbf{p})\rangle + \langle \mathbf{q}, \partial_p f(\mathbf{u},\mathbf{p})\rangle\,. \tag{26}$$

This can be done because all derivatives are by definition linear with respect to the differentiation direction and therefore the new expressions are still bilinear. In order to avoid overnotation we did not indicate that the bilinear forms $\langle\,.\,,\,.\,\rangle$ in (26) are not identical to the one in (25). Differentiating the operator equation (22) with respect to $\mathbf{p}$ yields

$$D_p\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{q}) = \partial_u\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{v}) + \partial_p\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{q}) = \mathbf{0}\,. \tag{27}$$

The dependence of $\mathbf{L}$ on $\mathbf{x}$ and t has been omitted in the notation. Now let $\boldsymbol{\psi}$ be an arbitrary but sufficiently nice test function. Since $\langle\,.\,,\,.\,\rangle$ is independently linear in both arguments we find

$$\langle \boldsymbol{\psi}, \partial_u\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{v})\rangle + \langle \boldsymbol{\psi}, \partial_p\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{q})\rangle = 0\,. \tag{28}$$

Assuming that there exist two operators $\partial_u\mathbf{L}^*(\mathbf{u};\mathbf{p})$ and $\partial_p\mathbf{L}^*(\mathbf{u};\mathbf{p})$ satisfying the relations

$$\langle \boldsymbol{\psi}, \partial_u\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{v})\rangle = \langle \mathbf{v}, \partial_u\mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi})\rangle \quad \text{and} \tag{29a}$$
$$\langle \boldsymbol{\psi}, \partial_p\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{q})\rangle = \langle \mathbf{q}, \partial_p\mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi})\rangle\,, \tag{29b}$$

equation (28) can be rewritten as

$$\langle \mathbf{v}, \partial_u\mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi})\rangle + \langle \mathbf{q}, \partial_p\mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi})\rangle = 0\,. \tag{30}$$

It is part of the problem to find the *transposed operators* $\partial_u\mathbf{L}^*(\mathbf{u};\mathbf{p})$ and $\partial_p\mathbf{L}^*(\mathbf{u};\mathbf{p})$. In the case of the two-dimensional scalar wave equation operator, integration by parts and Gauss' theorem lead to expressions for $\partial_u\mathbf{L}^*(\mathbf{u};\mathbf{p})$ and $\partial_p\mathbf{L}^*(\mathbf{u};\mathbf{p})$. But still, their existence was bound to a number of conditions on the test function $\boldsymbol{\psi}$. This means that the transposes exist only for a smaller class of test functions, namely those satisfying the adjoint subsidiary conditions. Adding equation (26) to equation (30) leads to

$$D_p\mathfrak{E}(\mathbf{u},\mathbf{p})(\mathbf{q}) = \langle \mathbf{v}, \partial_u f(\mathbf{u},\mathbf{p}) + \partial_u\mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi})\rangle + \langle \mathbf{q}, \partial_p f(\mathbf{u},\mathbf{p}) + \partial_p\mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi})\rangle\,. \tag{31}$$

The term $\mathbf{v} = D_p\mathbf{u}(\mathbf{p};\mathbf{x},t)$ may now be eliminated from (31) by imposing

$$\partial_u f(\mathbf{u},\mathbf{p}) + \partial_u\mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi}) = \mathbf{0}\,. \tag{32}$$

This is the *adjoint equation*. Together with the adjoint subsidiary conditions, required for the existence of the transposed operators, it forms the *adjoint problem*. Its solution $\boldsymbol{\psi}$, if it exists, is the *adjoint field*. If the adjoint problem can be solved, we can express $D_p\mathfrak{E}(\mathbf{u},\mathbf{p})(\mathbf{q})$ as

$$D_p\mathfrak{E}(\mathbf{u},\mathbf{p})(\mathbf{q}) = \langle \mathbf{q}, \partial_p f(\mathbf{u},\mathbf{p}) + \partial_p \mathbf{L}^*(\mathbf{u};\mathbf{p})(\boldsymbol{\psi})\rangle\,. \tag{33}$$

Equation (33) relates the total derivative of the objective function $\mathfrak{E}$ to a bilinear functional of the direction $\mathbf{q}$ and the adjoint field $\boldsymbol{\psi}$. Therefore, to obtain $D_p\mathfrak{E}(\mathbf{u},\mathbf{p})(\mathbf{q})$, it is sufficient to solve only once the original problem and the adjoint problem.
At this point let us briefly summarise the assumptions that we made so far. Firstly, we assumed that there exist two transposed operators. This may indeed be an obstacle. Depending on the operator $\mathbf{L}$, it may not be possible to find them. Secondly, we required an adjoint field ψ, satisfying both the adjoint subsidiary conditions and the adjoint equation. However, the adjoint problem may not have a solution, and if the solution exists, it may not be unique. Fortunately, we can find and uniquely solve the adjoint problems for all linear wave propagation phenomena, i.e., for all types of linear wave equation operators $\mathbf{L}$ complemented by adequate subsidiary conditions.

Bilinear operators: The theory outlined so far simplifies significantly if the operator $\mathbf{L}$ is bilinear, i.e., if it satisfies the relation

$$\mathbf{L}(\alpha\mathbf{u}+\beta\mathbf{v};\gamma\mathbf{p}+\varepsilon\mathbf{q}) = \alpha\gamma\mathbf{L}(\mathbf{u};\mathbf{p}) + \alpha\varepsilon\mathbf{L}(\mathbf{u};\mathbf{q}) + \beta\gamma\mathbf{L}(\mathbf{v};\mathbf{p}) + \beta\varepsilon\mathbf{L}(\mathbf{v};\mathbf{q}) \tag{34}$$

for all fields $\mathbf{u},\mathbf{v},\mathbf{p},\mathbf{q}$ and for all scalars $\alpha,\beta,\gamma,\varepsilon$. Due to the bilinearity, the total derivative of $\mathbf{L}(\mathbf{u};\mathbf{p})$ with respect to the parameters $\mathbf{p}$ is

$$D_p\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{q}) = \partial_u\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{v}) + \partial_p\mathbf{L}(\mathbf{u};\mathbf{p})(\mathbf{q}) = \mathbf{L}(\mathbf{v};\mathbf{p}) + \mathbf{L}(\mathbf{u};\mathbf{q})\,, \tag{35}$$

where $\mathbf{v}(\mathbf{p}) = D_p\mathbf{u}(\mathbf{p})(\mathbf{q})$. Introducing the notation $\partial_u\mathbf{L}^* = \mathbf{L}^u$ and $\partial_p\mathbf{L}^* = \mathbf{L}^p$, equations (32) and (33) reduce to

$$\mathbf{L}^u(\boldsymbol{\psi};\mathbf{p}) + \partial_u f = \mathbf{0}\,, \qquad D_p\mathfrak{E}(\mathbf{u};\mathbf{p})(\mathbf{q}) = \langle \mathbf{q}, \partial_p f(\mathbf{u};\mathbf{p}) + \mathbf{L}^p(\mathbf{u};\boldsymbol{\psi})\rangle\,. \tag{36}$$

It is straightforward to verify that the scalar wave equation operator defined in (23) is bilinear, i.e., independently linear in u and $\mathbf{p} = (\rho,\mu)$. Moreover, in this specific case, we find the identity

$$L^u(\psi;\mathbf{p}) = \rho\partial_t^2\psi - \nabla\cdot(\mu\,\nabla\psi) = L(\psi;\mathbf{p})\,. \tag{37}$$

Therefore, the two-dimensional scalar wave equation is *self-adjoint*. This property is closely related to energy conservation and spatial reciprocity. Note that the adjoint equation is independent of the field $\mathbf{u}$ in the case of a bilinear operator.

Non-linear operators: An important mathematical aspect of the adjoint

method is that we are strictly speaking not interested in the transpose of the operator $\mathbf{L}$ but in the transposes of its two partial derivatives $\partial_u \mathbf{L}$ and $\partial_p \mathbf{L}$. Since the derivatives are by definition linear in the differentiation directions we can hope to find transposes even if $\mathbf{L}$ itself is non-linear. As an example we analyse the substantial derivative term

$$\mathbf{L}(\mathbf{u}(\mathbf{p})) = (\mathbf{u} \cdot \nabla)\mathbf{u}\,, \tag{38}$$

which appears in the Navier-Stokes equations. It is assumed that $\mathbf{u}$ is defined in a region G and that the condition $\mathbf{u} \cdot d\mathbf{\Gamma} = 0$ holds for every boundary element $d\mathbf{\Gamma}$. Differentiating $\mathbf{L}$ with respect to $\mathbf{u}$ in the direction $\mathbf{v}$ yields

$$\langle \boldsymbol{\psi}, \partial_u \mathbf{L}(\mathbf{u})(\mathbf{v}) \rangle = \langle \boldsymbol{\psi}, (\mathbf{u} \cdot \nabla)\mathbf{v} \rangle + \langle \boldsymbol{\psi}, (\mathbf{v} \cdot \nabla)\mathbf{u} \rangle\,, \tag{39}$$

where $\boldsymbol{\psi}$ is a test function. The bilinear form is defined as $\langle\,.\,,\,.\,\rangle = \int_G (\,.\,,\,.\,) d^3\mathbf{x}$. The first summand in equation (39) can be transformed with Gauss' theorem,

$$\langle \boldsymbol{\psi}, (\mathbf{u} \cdot \nabla)\mathbf{v} \rangle = -\langle \mathbf{v}, (\mathbf{u} \cdot \nabla)\boldsymbol{\psi} \rangle - \langle \mathbf{v}, \boldsymbol{\psi}(\nabla \cdot \mathbf{u}) \rangle\,. \tag{40}$$

Combining equations (39) and (40) gives

$$\langle \boldsymbol{\psi}, \partial_u \mathbf{L}(\mathbf{u})(\mathbf{v}) \rangle = \langle \mathbf{v}, (\nabla \mathbf{u}) \cdot \boldsymbol{\psi} - \boldsymbol{\psi}(\nabla \cdot \mathbf{u}) - (\mathbf{u} \cdot \nabla)\boldsymbol{\psi} \rangle \tag{41}$$

and therefore

$$\partial_u \mathbf{L}^*(\mathbf{u}(\mathbf{p}))(\boldsymbol{\psi}) = (\nabla \mathbf{u}) \cdot \boldsymbol{\psi} - \boldsymbol{\psi}(\nabla \cdot \mathbf{u}) - (\mathbf{u} \cdot \nabla)\boldsymbol{\psi}\,. \tag{42}$$

This result may become important in fluid dynamics inverse problems that go beyond the Boussinesq approximation. Note that in this particular case the adjoint equations for $\boldsymbol{\psi}$ depend on the original field $\mathbf{u}$.

5 Objective functions

In this section we will consider some specific objective functions $\mathfrak{E}$ and the corresponding source terms $\partial_u f$ in the adjoint equations. The analysis will be based on the bilinear form

$$\langle \alpha, \beta \rangle := \int_{t=t_0}^{t_1} \int_G \alpha(\mathbf{x}, t)\beta(\mathbf{x}, t)\, dt\, d^n\mathbf{x}\,, \tag{43}$$

for two arbitrary integrable functions α and β and $G \subset \mathbb{R}^n$.

Jacobians: Let the objective function equal the i^{th} component of the vector field $\mathbf{u}$ at the point $\mathbf{x} = \boldsymbol{\xi}$ and the time $t = \tau$, i.e., $\mathfrak{E}(\mathbf{u}, \mathbf{p}) = \mathbf{e}_i \cdot \mathbf{u}(\boldsymbol{\xi}, \tau) =:$

$u_i(\boldsymbol{\xi}, \tau)$. The relation $\mathfrak{E}(\mathbf{u}, \mathbf{p}) = \langle 1, f(\mathbf{u}, \mathbf{p}; \mathbf{x}, t) \rangle$ implies

$$f(\mathbf{u}, \mathbf{p}; \mathbf{x}, t) = f(\mathbf{u}; \mathbf{x}, t) = \mathbf{e}_i \cdot \mathbf{u}(\mathbf{x}, t)\, \delta(\mathbf{x} - \boldsymbol{\xi})\, \delta(t - \tau)\,. \tag{44}$$

It follows that the source term of the adjoint equation is given by

$$\partial_u f = \mathbf{e}_i\, \delta(\mathbf{x} - \boldsymbol{\xi})\, \delta(t - \tau)\,. \tag{45}$$

This means that the adjoint field $\boldsymbol{\psi}$ has its source acting on a point at the observation time τ and the observation point $\boldsymbol{\xi}$. Given that the model space has a finite dimension m and basis vectors $\mathbf{e}_k$ with $k = 1, ..., m$, the derivative of $\mathfrak{E}$ with respect to the parameters $\mathbf{e}_k$ gives the Jacobian of $\mathbf{u}$, i.e.,

$$D_p \mathfrak{E}(\mathbf{u}, \mathbf{p})(\mathbf{e}_k) = D_p u_i(\mathbf{p})(\mathbf{e}_k) = \frac{\partial}{\partial p_k} u_i(\boldsymbol{\xi}, \tau)\,. \tag{46}$$

Generalised least squares: Let $\mathbf{u}^0(\boldsymbol{\xi}, \tau)$ denote values of the observable $\mathbf{u}$ measured over time t at the location $\mathbf{x} = \boldsymbol{\xi}$. When measurement and modelisation errors, as well as departures from the a priory model $\mathbf{p_0}$ can be modelled with Gaussian distributions, the objective function $\mathfrak{E}$ may be given by

$$\begin{aligned} \mathfrak{E}(\mathbf{u}, \mathbf{p}) =& \frac{1}{2} \int_{-\infty}^{\infty} [\mathbf{u}(\boldsymbol{\xi}, t) - \mathbf{u}^0(\boldsymbol{\xi}, t)] \cdot \mathbf{C}_d^{-1}(t) \cdot [\mathbf{u}(\boldsymbol{\xi}, t) - \mathbf{u}^0(\boldsymbol{\xi}, t)]\, dt \\ &+ \frac{1}{2} \int_G [\mathbf{p}(\mathbf{x}) - \mathbf{p}_0(\mathbf{x})] \cdot \mathbf{C}_p^{-1}(\mathbf{x}) \cdot [\mathbf{p}(\mathbf{x}) - \mathbf{p}_0(\mathbf{x})]\, d^3\mathbf{x}\,. \end{aligned} \tag{47}$$

The symmetric second order tensor $\mathbf{C}_d^{-1}$ describes measurement and modelisation errors, while the symmetric second order tensor $\mathbf{C}_p^{-1}$ contains all prior information on the parameter space (e.g. Tarantola, 1987). The function f corresponding to $\mathfrak{E}$ is

$$\begin{aligned} f(\mathbf{u}; \mathbf{x}, t) =& \frac{1}{2} [\mathbf{u}(\mathbf{x}, t) - \mathbf{u}^0(\mathbf{x}, t)] \cdot \mathbf{C}_d^{-1}(t) \cdot [\mathbf{u}(\mathbf{x}, t) - \mathbf{u}^0(\mathbf{x}, t)]\, \delta(\mathbf{x} - \boldsymbol{\xi}) \\ &+ \frac{1}{2} [\mathbf{p}(\mathbf{x}) - \mathbf{p}_0(\mathbf{x})] \cdot \mathbf{C}_p^{-1}(\mathbf{x}) \cdot [\mathbf{p}(\mathbf{x}) - \mathbf{p}_0(\mathbf{x})]\, \delta(t)\,. \end{aligned} \tag{48}$$

This translates to the following source term of the adjoint problem:

$$\partial_u f = \mathbf{C}_d^{-1}(t) \cdot [\mathbf{u}(\mathbf{x}, t) - \mathbf{u}^0(\mathbf{x}, t)]\, \delta(\mathbf{x} - \boldsymbol{\xi})\,. \tag{49}$$

A point source is located directly at the observation point $\mathbf{x} = \boldsymbol{\xi}$. Its source time function and the amplitude of the source are specified by the development of the residuals $\mathbf{u} - \mathbf{u}_0$ over time, weighted by $\mathbf{C}_d^{-1}$.

6 The anisotropic wave equation with attenuation

As an application of the theory outlined so far we analyse in this section the adjoint problem for the anisotropic wave equation with attenuation. Similar results have already been obtained by Tarantola (1988) who used an approach based on the Born approximation and the existence of the Green's function. The linearised conservation of momentum is given by

$$\rho(\mathbf{x})\partial_t^2\mathbf{u}(\mathbf{x},t) - \nabla\cdot\mathbf{T}(\mathbf{x},t) = \mathbf{g}(\mathbf{x},t)\,, \tag{50}$$

where ρ is the initial mass density distribution, $\mathbf{u}$ the incremental displacement field, $\mathbf{T}$ the incremental stress tensor and $\mathbf{g}$ a body force density. The incremental stress tensor $\mathbf{T}$ can be related to the displacement field $\mathbf{u}$ via the constitutive relation

$$\mathbf{T}(\mathbf{x},t) = \int_{-\infty}^{\infty} \mathbf{\Phi}(\mathbf{x},t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau\,. \tag{51}$$

The rate of relaxation function $\mathbf{\Phi}$ is a fourth order tensor assumed to satisfy the relations

$$\mathbf{\Phi}(\mathbf{x},t)|_{t<0} = 0 \quad \text{and} \quad \Phi_{ijkl} = \Phi_{klij} = \Phi_{jikl}\,, \tag{52}$$

i.e., causality and elastic symmetry. In the case of cartesian coordinates the invariant notation $\mathbf{\Phi} : \nabla\mathbf{u}$ is equivalent to $\Phi_{ijkl}\partial_k u_l$ in index notation. Inserting (51) into (50) allows us to express the wave equation operator $\mathbf{L}$ in terms of the displacement field $\mathbf{u}$ and the medium parameters $\mathbf{\Phi}$ and ρ,

$$\mathbf{L}(\mathbf{u};\mathbf{\Phi},\rho,\mathbf{x},t) = \rho(\mathbf{x})\partial_t^2\mathbf{u}(\mathbf{x},t) - \nabla\cdot\int_{-\infty}^{t} \mathbf{\Phi}(\mathbf{x},t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau\,, \tag{53}$$

for $t \in [t_0, t_1]$. In order to guarantee uniqueness, the equations have to be complemented by a set of subsidiary conditions.

$$\mathbf{u}(\mathbf{x},t)|_{t=t_0} = \mathbf{g}_1(\mathbf{x})\,, \qquad \text{(initial cond.)} \tag{54a}$$

$$\partial_t\mathbf{u}(\mathbf{x},t)|_{t=t_0} = \mathbf{g}_2(x)\,, \qquad \text{(initial cond.)} \tag{54b}$$

$$\mathbf{u}(\mathbf{x},t)|_{t<t_0} = \mathbf{0}\,, \qquad \text{(quiescent past)} \tag{54c}$$

$$\mathbf{u}(\mathbf{x},t)|_{\mathbf{x}\in\Gamma_1} = \mathbf{b}_1(\mathbf{x},t)\,, \qquad \text{(Cauchy cond.)} \tag{54d}$$

$$\mathbf{n}\cdot\int_{-\infty}^{t} \mathbf{\Phi}(\mathbf{x},t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau\Big|_{\mathbf{x}\in\Gamma_2} = \mathbf{b}_2(\mathbf{x},t)\,, \qquad \text{(Neumann cond.)} \tag{54e}$$

Again, we require $\partial G = \Gamma_1 \cup \Gamma_2$. It is convenient to decompose $\mathbf{L}$ into $\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$ with

$$\mathbf{L}_1(\mathbf{u};\rho,\mathbf{x},t) := \rho(\mathbf{x})\partial_t^2\mathbf{u}(\mathbf{x},t)\,, \tag{55}$$

$$\mathbf{L}_2(\mathbf{u};\mathbf{\Phi},\mathbf{x},t) := -\nabla\cdot\int_{-\infty}^{t}\mathbf{\Phi}(\mathbf{x},t-\tau):\nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau\,, \tag{56}$$

and to analyse the two summands separately. Using the bilinear form

$$\langle\mathbf{a},\mathbf{b}\rangle := \int_G\int_{t_0}^{t_1}\mathbf{a}(\mathbf{x},t)\cdot\mathbf{b}(\mathbf{x},t)\,d^3\mathbf{x}\,dt\,, \tag{57}$$

for any two integrable fields $\mathbf{a}$ and $\mathbf{b}$, we find

$$\begin{aligned}\langle\boldsymbol{\psi},D_\rho\mathbf{L}_1(\mathbf{u};\rho)(\rho')\rangle =&\langle\boldsymbol{\psi},\partial_\rho\mathbf{L}_1(\mathbf{u};\rho)(\rho')\rangle + \langle\boldsymbol{\psi},\partial_u\mathbf{L}_1(\mathbf{u};\rho)(\mathbf{v}_1)\rangle\\ =&\langle\boldsymbol{\psi},\rho'\partial_t^2\mathbf{u}\rangle + \langle\boldsymbol{\psi},\rho\partial_t^2\mathbf{v_1}\rangle\\ =&\langle\rho',\boldsymbol{\psi}\cdot\partial_t^2\mathbf{u}\rangle + \langle\mathbf{v_1},\rho\partial_t^2\boldsymbol{\psi}\rangle\\ =&\langle\rho',\partial_\rho\mathbf{L}_1^*(\mathbf{u};\rho)(\boldsymbol{\psi})\rangle + \langle\mathbf{v}_1,\partial_u\mathbf{L}_1^*(\mathbf{u};\rho)(\boldsymbol{\psi})\rangle = 0\,,\end{aligned} \tag{58}$$

with $\mathbf{v}_1(\mathbf{\Phi},\rho;\mathbf{x},t) := D_\rho\mathbf{u}(\mathbf{\Phi},\rho;\mathbf{x},t)(\rho')$. To obtain this equality, we repeatedly integrated by parts and imposed the *homogeneous* terminal conditions $\boldsymbol{\psi}(\mathbf{x},t)|_{t=t_1} = \mathbf{0}$ and $\partial_t\boldsymbol{\psi}(\mathbf{x},t)|_{t=t_1} = \mathbf{0}$ onto the adjoint field. It remains to consider the spatial derivative operator $\mathbf{L}_2$. Due to the bilinearity of $\mathbf{L}_2$ with respect to $\mathbf{\Phi}$ and $\mathbf{u}$, its derivative with respect to $\mathbf{\Phi}$ in the direction of $\mathbf{\Phi'}$ is

$$\begin{aligned}D_\Phi\mathbf{L}_2(\mathbf{u};\mathbf{\Phi},\mathbf{x},t)(\mathbf{\Phi'}) &= \partial_\Phi\mathbf{L}_2(\mathbf{u};\mathbf{\Phi},\mathbf{x},t)(\mathbf{\Phi'}) + \partial_u\mathbf{L}_2(\mathbf{u};\mathbf{\Phi},\mathbf{x},t)(\mathbf{v}_2)\\ &= -\nabla\cdot\int_{-\infty}^{t}\mathbf{\Phi}(\mathbf{x},t-\tau):\nabla\mathbf{v_2}(\mathbf{x},\tau)\,d\tau\\ &\quad-\nabla\cdot\int_{-\infty}^{t}\mathbf{\Phi'}(\mathbf{x},t-\tau):\nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau\,,\end{aligned} \tag{59}$$

with $\mathbf{v}_2(\mathbf{\Phi},\rho;\mathbf{x},t) := D_\Phi\mathbf{u}(\mathbf{\Phi},\rho;\mathbf{x},t)(\mathbf{\Phi'})$. The perturbation or differentiation direction $\mathbf{\Phi'}$ should satisfy the same symmetry relations as $\mathbf{\Phi}$ and be causal. In the appendix, we demonstrate that the complete adjoint problem is

$$\mathbf{L}^u(\boldsymbol{\psi};\mathbf{\Phi},\rho,\mathbf{x},t) = \rho(\mathbf{x})\partial_t^2\boldsymbol{\psi}(\mathbf{x},t) - \nabla\cdot\int_{-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},\tau-t):\nabla\boldsymbol{\psi}(\mathbf{x},\tau)\,d\tau = -\partial_u f\,, \tag{60}$$

with subsidiary conditions

$$\boldsymbol{\psi}(\mathbf{x},t)|_{t=t_1} = \mathbf{0}\,, \qquad \text{(terminal condition)} \tag{61a}$$

$$\partial_t\boldsymbol{\psi}(\mathbf{x},t)|_{t=t_1} = \mathbf{0}\,, \qquad \text{(terminal condition)} \tag{61b}$$

$$\boldsymbol{\psi}(\mathbf{x},t)|_{t>t_1} = \mathbf{0}\,, \qquad \text{(quiescent future)} \tag{61c}$$

$$\boldsymbol{\psi}(\mathbf{x},t)|_{\mathbf{x}\in\Gamma_1} = \mathbf{0}\,, \qquad \text{(Cauchy condition)} \tag{61d}$$

$$\mathbf{n}\cdot\int_{-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},\tau-t):\nabla\boldsymbol{\psi}(\mathbf{x},\tau)\,d\tau\Big|_{\mathbf{x}\in\Gamma_2} = \mathbf{0}\,. \qquad \text{(Neumann condition)} \tag{61e}$$

The explicit formula for the derivative of the objective function $\mathfrak{E}$ is

$$
\begin{aligned}
D_{(\Phi,\rho)}\mathfrak{E}(\mathbf{u},\boldsymbol{\Phi},\rho)((\boldsymbol{\Phi}',\rho')) &= D_\rho\mathfrak{E}(\mathbf{u},\boldsymbol{\Phi},\rho)(\rho') + D_\Phi\mathfrak{E}(\mathbf{u},\boldsymbol{\Phi},\rho)(\boldsymbol{\Phi}') \\
&= \int_G \int_{t=t_0}^{t_1} \rho'(\mathbf{x}) \left[\partial_\rho f + \boldsymbol{\psi}(\mathbf{x},t)\cdot\partial_t^2\mathbf{u}(\mathbf{x},t)\right] dt\, d^3\mathbf{x} \\
&+ \int_G \int_{t=0}^{t_1-t_0} \boldsymbol{\Phi}'(\mathbf{x},t) :: \left[\partial_\Phi f + \int_{\tau=t_0+t}^{t_1} \nabla\boldsymbol{\psi}(\mathbf{x},\tau)\otimes\nabla\mathbf{u}(\mathbf{x},\tau-t)\, d\tau\right] dt\, d^3\mathbf{x}\,.
\end{aligned}
\tag{62}
$$

The symbol $\otimes$ denotes the tensor or dyadic product ($(\mathbf{a}\otimes\mathbf{b})_{ijkl} = a_{ij}b_{kl}$) and the symbol :: the quadruple scalar product ($\mathbf{A} :: \mathbf{B} = A_{ijkl}B_{ijkl}$). In these formulae we prefer the invariant tensor notation because it is better suited to reveal the structure of an equation and moreover, it is valid in all coordinate systems.

The most remarkable similarity between the adjoint problem and the original problem is the structure of the equations themselves. They are linear and the time derivative term translates one to one from the original to the adjoint equation. The invariance of the time derivative term is due to the invariance of a second derivative to a sign change. This symmetry is closely related to the conservation of energy. It is therefore not surprising that the spatial derivative term is different in the adjoint equations because it incorporates the loss of elastic energy in the form of a time-dependent rheology. Interestingly, the sign change in the temporal variable t provokes a transition from *causality* to *anti-causality*. The time reversal in the governing equations also translates to a time reversal in the respective subsidiary conditions, meaning that the adjoint field $\boldsymbol{\psi}$ is required to have a *quiescent future* (instead of a quiescent past, as $\mathbf{u}$) and thus *no future strains* that could translate to *present stresses*. Also, note that the spatial subsidiary conditions of the adjoint problem, i.e., the adjoint Cauchy and Neumann conditions, are necessarily *homogeneous*, meaning that the boundaries are either free (no stresses) or rigid (no displacements). The terminal conditions are *homogeneous* as well, irrespective of possible non-zero initial displacement or velocity distributions.

Equations (60) to (62) were obtained by Tarantola (1988). He founded his derivation on the first-order Born approximation and the reciprocity relations of Green's functions. These equations are in a certain sense very optimistic. In most applications it is not possible to specify spatial variations of anelasticity or to determine complete rate of relaxation functions. Simplified equations for perfectly elastic and isotropic media can in general be obtained by specifying the rate of relaxation tensor $\boldsymbol{\Phi}$.

7 Discussion

The first derivative of a physical observable with respect to the parameters determining its properties is of major interest in all quantitative sciences. Large parameter spaces as those of modern Earth models usually render a direct approximation of that derivative by means of finite differencing impossible. An elegant and efficient solution for this problem is the adjoint method introduced into the field of seismology by Tarantola (1984, 1988). The adjoint method allows us to compute the derivative of an objective function with respect to the model parameters by simply solving the original problem and the adjoint problem only once.

In this paper we first analysed the adjoint problem for the two-dimensional scalar wave equation and the least squares objective function. This simple example served to illustrate both the methodology and the physical meaning of the adjoint wavefield. In principle, the adjoint method is a mathematical technique that allows us to eliminate the unknown derivative of the observed wavefield from the derivative of the objective function by introducing a new variable ψ.

The two-dimensional scalar wave equation coincides with its adjoint equation. Also the boundary conditions translate one-to-one from the original to the adjoint problem, whereas the initial conditions translate to terminal conditions. This property of the adjoint problem forces us to solve it by stepping backwards in time. For the computation of the drivative $D_p\mathfrak{E}$, the forward wavefield and the adjoint wavefield have to be known at the same point in time. When the forward problem is not dissipative, $D_p\mathfrak{E}$ can be computed very efficiently without storing the forward wavefield u. This is because u can in principle be propagated backward in time, i.e., from t_1 to t_0, starting with its known final state $u(\mathbf{x}, t_1)$. The adjoint field ψ is then being computed simultaneously and also from t_1 to t_0. In the case of a dissipative forward problem it becomes technically infeasible to propagate the forward field backward in time because lost information can hardly be recovered. Hence, the forward field must be stored, at least for a certain number of time steps. In this context dissipation must be seen in a broader sense. It includes physical and numerical dissipation but also absorbing boundary conditions implemented through damping regions (e.g. Cerjan et al., 1985).

The source term of the adjoint equation only depends on the objective function. It is in particular independent of the sources of the forward wavefield. This offers the possibility to define the forward wavefield as a superposition of wavefields corresponding to different sources acting at conveniently chosen points in time. In the case of the least squares objective function the adjoint source is restricted to the receiver locations and its source time function is determined by the time evolution of the linear residuals. In a simple numerical example we demonstrated that the adjoint wavefield focusses at the location of a parameter perturbation. Such a perturbation may for example be a dif-

ference between the true Earth model and the numerical model. When this focussing occurs the wavefront of the original field also reaches the location of the parameter perturbation. This coincidence leads to a contribution to the derivative $D_p\mathfrak{E}$. An implicaton for any gradient-method based waveform inversion is that one has to account for the anisotropic structure of the Earth. Neglecting anisotropy would lead to significant problems because the adjoint field may focus in the wrong positions or it may not focus at all.
Using a condensed notation we found an elegant generalisation of the adjoint method. The employed terminology is certainly not common in the geophysical literature; it is however modern mathematical standard and therefore also widely used in theoretical physics. We saw that the adjoint method relies on the existence of transposes of the partial derivatives of an operator. Since a derivative is by definition linear in the differentiation direction, the adjoint method becomes applicable to non-linear equations such as the Navier-Stokes equations. In the presence of non-linearity the adjoint field depends directly on the forward field. For example, the adjoint of the substantial derivative term involves products of the forward velocity field and the adjoint field. Moreover, the resulting adjoint equation will not be similar to the orginal equation involving the substantial derivative term. The numerical effort needed to solve the adjoint problem is therefore significantly increased. Sill, the non-linear adjoints may become important in future geodynamic applications that go beyond the usual linearisations.
As our final application we considered the elastic wave equation with attenuation and anisotropy. This forward problem is certainly to general for most seismological applications. However, it provides interesting insight into the physics of the problem. Also, the resulting equations can easily be simplified. The wave equation operator looses its symmetry due to the presence of anelasticity. Hence, the adjoint problem differs from the original problem.

8 Conclusions

The adjoint method is an elegant and efficient tool for the computation of the first derivative of an objective function with respect to model parameters. Using modern mathematical terminology one finds that the adjoint method produces exact derivatives and that it does not rely on the existence of Green's functions or transposes of a differential operator. Only the transposes of its partial derivatives are of interest. The method can be applied to non-linear operators such as the one corresponding to the Navier-Stokes equations. In the case of the wave equation and the least squares objective function the adjoint field focusses at the locations of parameter perturbations. The simultaneous passage of the original wavefront leads to a gradient contribution. This im-

plies that gradient-based waveform inversion procedures should account for anisotropy.

Acknowledgements: The authors are grateful to Brian Kennett, Steve Haney, Bernhard Forkmann and Wolfgang Sproessig for interesting discussions and for providing support and constructive criticism. The comments of several anonymous reviewers helped to improve the manuscript.

References

[1] Baysal, E., 1983. Reverse time migration, *Geophysics*, **48**, No. 11, 1514-1524.

[2] Becker, T. W., L. Boschi, 2002. A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, **3**, 10.129/2001GC000168.

[3] Brown, J. M., T. J. Shankland, 1981. Thermodynamic parameters in the Earth as determined from seismic profiles, *Geophys. J. R. Astr. Soc.*, **66**, 579-596.

[4] Bunge, H.-P., M. A. Richards, C. Lithgow-Bertelloni, J. R. Baumgardner, S. Grand, B. Romanowicz, 1998. Time scales and heterogeneous structure in geodynamic earth models, *Science*, **280**, 91-95

[5] Bunge, H.-P., M. A. Richards, J. R. Baumgardner, 2002. Mantle circulation models with sequential data-assimilation: Inferring present-day mantle structure from plate motion histories, *Phil. Trans. Roy. Soc. A*, **360** (1800), 2545-2567.

[6] Bunge, H.-P., C. R. Hagelberg, B. J. Travis, 2003. Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography, *Geophys. J. Int.*, **152**, 280-301.

[7] Cerjan, C., D. Kosloff, R. Kosloff, M. Reshef, 1985. A non-reflecting boundary condition for discrete acoustic and elastic wave equations, *Geophysics*, **50**, 705-708.

[8] Crase, I., A. Pica, M. Noble, J. McDonald, A. Tarantola, 1990. Robust elastic nonlinear waveform inversion: Application to real data, *Geophysics*, **55**(5), 527-538.

[9] Duffy, T. S., T. J. Ahrens, 1992. Sound velocities at high pressure and temperature and their geophysical implications, *J. Geophys. Res.*, **97**, 4503-4520.

[10] Dziewonski, A. M., A. L. Hales, E. R. Lapwood, 1975. Parametrically simple Earth models consistent with geophysical data, *Phys. Earth Planet. Int.*, **10**, 12-48.

[11] Dziewonski, A. M., B. H. Hager, R. J. O'Connell, 1977. Large-scale heterogeneities in the lower mantle, *J. Geophys. Res.*, **82**(2), 239-255.

[12] Dziewonski, A. M., D. L. Anderson, 1981. Preliminary reference Earth model, *Phys. Earth Planet. Int.*, **25**, 297-356.

[13] Dziewonski, A. M., 1984. Mapping the lower mantle: Determination of lateral heterogeneity in P-velocity up to degree and order 6, *J. Geophys. Res.*, **89**, 5929-5952.

[14] Gauthier, O., J. Virieux, A. Tarantola, 1986. Two-dimensional nonlinear inversion of seismic waveforms: Numerical results, *Geophysics*, **51**, No. 7, 1387-1403.

[15] Gorbatov, A., B. L. N. Kennett, 2003. Joint bulk-sound and shear tomography for Western Pacific subduction zones, *Earth Planet. Sci. Lett.*, **210**, 527-543.

[16] Grand, S. P., 1994. Mantle shear structure beneath the Americas and surrounding oceans, *J. Geophys. Res.*, **99**, 11,591-11,621.

[17] Grand, S., R. D. van der Hilst, S. Widiyantoro, 1997. Global seismic tomography: A snapshot of convection in the Earth, *GSA Today*, **7**, n. 4, 1-7.

[18] Igel, H., P. Mora, B. Riollet, 1995. Anisotropic wave propagation through finite-difference grids, *Geophysics*, **60**(4), 1203-1216.

[19] Igel, H., H. Djikpesse, A. Tarantola, 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, *Geophys. J. Int.*, **124**(2), 363-371.

[20] Kantorowitsch, L. W., G. P. Akilow, 1964. *Funktionalanalysis in normierten Räumen*, Akademie-Verlag, Berlin.

[21] Kárason, H., R. D. van der Hilst, 2000. Constraints on mantle convection from seismic tomography, in *The History and Dynamics of Global Plate Motions*, 277-288.

[22] Kárason, H., R. D. van der Hilst, 2001. Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP, P_{diff}), *J. Geophys. Res.*, **106**(B4), 6569-6587.

[23] Kennett, B. L. N., E. R. Engdahl, 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429-465.

[24] Kennett, B. L. N., E. R. Engdahl, R. Buland, 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, **122**, 108-124.

[25] Kennett, B. L. N., A. Gorbatov, 2004. Seismic heterogeneity in the mantle - strong shear wave signature of slabs from joint tomography, *Phys. Earth Planet. Int.*, **146**, 87-100.

[26] Komatitsch, D., C. Barnes, J. Tromp, 2000. Simulation of anisotropic wave propagation based upon a spectral element method, *Geophysics*, **65**(4), 1251-1260.

[27] Lithgow-Bertelloni, C., M. A. Richards, 1998. The dynamics of Cenozoic and Mesozoic plate motions, *Reviews of Geophysics*, **36**(1), 27-78.

[28] Masters, G., S. Johnson, G. Laske, H. Bolton, 1996. A shear-velocity model of the mantle, *Phil. Trans. R. Soc. Lond., A*, **354**, 1385-1411.

[29] Mattern, E., J. Matas, Y. Ricard, J. Bass, 2005. Lower mantle composition and temperature from mineral physics and thermodynamic modelling,

Geophys. J. Int., **160**(3), 973-990.
[30] Mégnin, C., B. Romanowicz, 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophys. J. Int.*, **143**, 709-728.
[31] Press, F., 1968. Earth models obtained by Monte Carlo inversion, *J. Geophys. Res.*, **73**, 5223-5234.
[32] Richards, M. A., D. C. Engebretson, 1992. Large-scale mantle convection and the history of subduction, *Nature*, 355, 437-440.
[33] Ritsema, J., H. J. van Heijst, 2000. Seismic imaging of structural heterogeneity in Earth's mantle: Evidence for large-scale mantle flow, *Sci. Progr.*, **83**, 243-259.
[34] Talagrand, O., P. Courtier, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, **113**, 1311-1328.
[35] Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, No. 8, 1259-1266.
[37] Tarantola, A., 1987. *Inverse problem theory. Methods for data fitting and model parameter estimation.*, Elsevier, Amsterdam.
[37] Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including anelasticity and attenuation, *Pure and Applied Geophysics*, **128**, 365-399.
[38] Tromp, J., C. Tape, Q. Liu, 2005. Seismic tomography, adjoint methods, time reversal, and banana-donut kernels, *Geophys. J. Int.*, **160**, 195-216.
[39] van der Hilst, R. D., S. Widiyantoro, E. R. Engdahl, 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**, 578-584.

A The adjoint equations for the anelastic wave equation with attenuation

a) Subsidiary conditions

Subsidiary conditions for the field $D_\Phi \mathbf{u}(\mathbf{x},t)(\mathbf{\Phi'}) =: \mathbf{v}_2(\mathbf{x},t)$ can be found through a simple application of the operator D_Φ to equations (54a) to (54e): $\mathbf{v}_2(\mathbf{x},t)|_{t=t_0} = \mathbf{0}$, $\partial_t^2 \mathbf{v}_2(\mathbf{x},t)|_{t=t_0} = \mathbf{0}$, $\mathbf{v}_2(\mathbf{x},t)|_{t<t_0} = \mathbf{0}$, $\mathbf{v}_2(\mathbf{x},t)|_{\mathbf{x}\in\Gamma_1} = \mathbf{0}$ and

$$\mathbf{n} \cdot \left[\int_{-\infty}^{\infty} \mathbf{\Phi}(\mathbf{x}, t-\tau) : \nabla \mathbf{v}_2(\mathbf{x},\tau)\, d\tau + \int_{-\infty}^{\infty} \mathbf{\Phi'}(\mathbf{x}, t-\tau) : \nabla \mathbf{u}(\mathbf{x},\tau)\, d\tau \right] \Bigg|_{\mathbf{x}\in\Gamma_2} = \mathbf{0}\,.$$

b) Displacement transpose

We already found

$$D_\Phi \mathbf{L}_2(\mathbf{u};\mathbf{\Phi},x,t)(\mathbf{\Phi'}) = \partial_\Phi \mathbf{L}_2(\mathbf{u};\mathbf{\Phi},x,t)(\mathbf{\Phi'}) + \partial_u \mathbf{L}_2(\mathbf{u};\mathbf{\Phi},\mathbf{x},t)(\mathbf{v}_2)\,, \quad \text{(A.1)}$$

with

$$\partial_\Phi \mathbf{L}_2(\mathbf{u};\mathbf{\Phi},x,t)(\mathbf{\Phi'}) = -\nabla\cdot\int\limits_{-\infty}^{t}\mathbf{\Phi'}(\mathbf{x},t-\tau):\nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau \tag{A.2}$$

and

$$\partial_u \mathbf{L}_2(\mathbf{u};\mathbf{\Phi},\mathbf{x},t)(\mathbf{v}_2) = -\nabla\cdot\int\limits_{-\infty}^{t}\mathbf{\Phi}(\mathbf{x},t-\tau):\nabla\mathbf{v}_2(\mathbf{x},\tau)\,d\tau\,. \tag{A.3}$$

First, we consider the difference

$$\begin{aligned} J_1 = &-\int\limits_G\int\limits_{t=t_0}^{t_1}\boldsymbol{\psi}(\mathbf{x},t)\cdot\left[\nabla\cdot\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},t-\tau):\nabla\mathbf{v}_2(\mathbf{x},\tau)\,d\tau\right]d^3\mathbf{x}\,dt\\ &+\int\limits_G\int\limits_{t=t_0}^{t_1}\mathbf{v}_2(\mathbf{x},t)\cdot\left[\nabla\cdot\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},\tau-t):\nabla\boldsymbol{\psi}(\mathbf{x},\tau)\,d\tau\right]d^3\mathbf{x}\,dt\,, \end{aligned} \tag{A.4}$$

for a sufficiently nice vector function $\boldsymbol{\psi}$. An application of Gauss' integral theorem directly yields $J_1 = J_{11} + J_{12}$, with

$$\begin{aligned} J_{11} :=& -\int\limits_G\int\limits_{t=t_0}^{t_1}\nabla\cdot\left[\boldsymbol{\psi}(\mathbf{x},t)\cdot\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},t-\tau):\nabla\mathbf{v}_2(\mathbf{x},\tau)\,d\tau\right]d^3\mathbf{x}\,dt\\ &+\int\limits_G\int\limits_{t=t_0}^{t_1}\nabla\cdot\left[\mathbf{v}_2(\mathbf{x},t)\cdot\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},\tau-t):\nabla\boldsymbol{\psi}(\mathbf{x},\tau)\,d\tau\right]d^3\mathbf{x}\,dt\\ =&-\int\limits_{\partial G}\int\limits_{t=t_0}^{t_1}\left[\boldsymbol{\psi}(\mathbf{x},t)\cdot\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},t-\tau):\nabla\mathbf{v}_2(\mathbf{x},\tau)\,d\tau\right]\cdot d\mathbf{\Gamma}\,dt\\ &+\int\limits_{\partial G}\int\limits_{t=t_0}^{t_1}\left[\mathbf{v}_2(\mathbf{x},t)\cdot\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},\tau-t):\nabla\boldsymbol{\psi}(\mathbf{x},\tau)\,d\tau\right]\cdot d\mathbf{\Gamma}\,dt \end{aligned} \tag{A.5}$$

and

$$\begin{aligned} J_{12} :=& \int\limits_G\int\limits_{t=t_0}^{t_1}[\nabla\boldsymbol{\psi}(\mathbf{x},t)]:\left[\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},t-\tau):\nabla\mathbf{v}_2(\mathbf{x},\tau)\,d\tau\right]d^3\mathbf{x}\,dt\\ &-\int\limits_G\int\limits_{t=t_0}^{t_1}[\nabla\mathbf{v}_2(\mathbf{x},t)]:\left[\int\limits_{\tau=-\infty}^{\infty}\mathbf{\Phi}(\mathbf{x},\tau-t):\nabla\boldsymbol{\psi}(\mathbf{x},\tau)\,d\tau\right]d^3\mathbf{x}\,dt\,. \end{aligned} \tag{A.6}$$

The causality property of $\mathbf{\Phi}$, i.e., the requirement that $\mathbf{\Phi}(t) = 0$ for $t < 0$, implies

$$
\begin{aligned}
J_{12} = &\int_G \int_{t=t_0}^{t_1} [\nabla \boldsymbol{\psi}(\mathbf{x},t)] : \left[\left(\int_{\tau=-\infty}^{t_0} + \int_{\tau=t_0}^{t_1} \right) \mathbf{\Phi}(\mathbf{x}, t-\tau) : \nabla \mathbf{v}_2(\mathbf{x},\tau)\, d\tau \right] d^3\mathbf{x}\, dt \\
&- \int_G \int_{t=t_0}^{t_1} [\nabla \mathbf{v}_2(\mathbf{x},t)] : \left[\left(\int_{\tau=t_0}^{t_1} + \int_{\tau=t_1}^{\infty} \right) \mathbf{\Phi}(\mathbf{x}, \tau-t) : \nabla \boldsymbol{\psi}(\mathbf{x},\tau)\, d\tau \right] d^3\mathbf{x}\, dt \\
= &\int_G \int_{t=t_0}^{t_1} \int_{\tau=-\infty}^{t_0} [\nabla \boldsymbol{\psi}(\mathbf{x},t)] : [\mathbf{\Phi}(\mathbf{x}, t-\tau) : \nabla \mathbf{v}_2(\mathbf{x},\tau)]\; d^3\mathbf{x}\, d\tau\, dt \\
&- \int_G \int_{t=t_0}^{t_1} \int_{\tau=t_1}^{\infty} [\nabla \mathbf{v}_2(\mathbf{x},t)] : [\mathbf{\Phi}(\mathbf{x}, \tau-t) : \nabla \mathbf{\Psi}(\mathbf{x},\tau)]\; d^3\mathbf{x}\, d\tau\, dt \\
&+ \int_G \int_{t=t_0}^{t_1} \int_{\tau=t_0}^{t_1} [\nabla \boldsymbol{\psi}(\mathbf{x},t)] : [\mathbf{\Phi}(\mathbf{x}, t-\tau) : \nabla \mathbf{v}_2(\mathbf{x},\tau)]\; d^3\mathbf{x}\, d\tau\, dt \\
&- \int_G \int_{t=t_0}^{t_1} \int_{\tau=t_0}^{t_1} [\nabla \mathbf{v}_2(\mathbf{x},t)] : [\mathbf{\Phi}(\mathbf{x}, \tau-t) : \nabla \boldsymbol{\psi}(\mathbf{x},\tau)]\; d^3\mathbf{x}\, d\tau\, dt\,. \qquad \text{(A.7)}
\end{aligned}
$$

Due to the symmetry of $\mathbf{\Phi}$, the last two summands of equation (A.7) cancel, and J_1 reduces to

$$
\begin{aligned}
J_1 = &- \int_{\partial G} \int_{t=t_0}^{t_1} \left[\boldsymbol{\psi}(\mathbf{x},t) \cdot \int_{\tau=-\infty}^{\infty} \mathbf{\Phi}(\mathbf{x}, t-\tau) : \nabla \mathbf{v}_2(\mathbf{x},\tau)\, d\tau \right] \cdot d\mathbf{\Gamma}\, dt \\
&+ \int_{\partial G} \int_{t=t_0}^{t_1} \left[\mathbf{v}_2(\mathbf{x},t) \cdot \int_{\tau=-\infty}^{\infty} \mathbf{\Phi}(\mathbf{x}, \tau-t) : \nabla \boldsymbol{\psi}(\mathbf{x},\tau)\, d\tau \right] \cdot d\mathbf{\Gamma}\, dt \\
&+ \int_G \int_{t=t_0}^{t_1} \int_{\tau=-\infty}^{t_0} [\nabla \boldsymbol{\psi}(\mathbf{x},t)] : [\mathbf{\Phi}(\mathbf{x}, t-\tau) : \nabla \mathbf{v}_2(\mathbf{x},\tau)]\; d^3\mathbf{x}\, d\tau\, dt \\
&- \int_G \int_{t=t_0}^{t_1} \int_{\tau=t_1}^{\infty} [\nabla \mathbf{v}_2(\mathbf{x},t)] : [\mathbf{\Phi}(\mathbf{x}, \tau-t) : \nabla \boldsymbol{\psi}(\mathbf{x},\tau)]\; d^3\mathbf{x}\, d\tau\, dt\,. \qquad \text{(A.8)}
\end{aligned}
$$

c) Rate of relaxation transpose

We are now interested in transforming the expression

$$
J_2 + I_2 = - \int_G \int_{t=t_0}^{t_1} \boldsymbol{\psi}(\mathbf{x},t) \cdot \left[\int_{\tau=-\infty}^{\infty} \mathbf{\Phi'}(\mathbf{x}, t-\tau) : \nabla \mathbf{u}(\mathbf{x},\tau)\, d\tau \right] dt\, d^3\mathbf{x}\,, \quad \text{(A.9)}
$$

Fig. A.1. The shaded area corresponds to the temporal integration domain Ω_T.

with

$$J_2 := -\int\limits_G \int\limits_{t=t_0}^{t_1} \nabla \cdot \left[\boldsymbol{\psi}(\mathbf{x},t) \cdot \int\limits_{\tau=-\infty}^{\infty} \boldsymbol{\Phi'}(\mathbf{x},t-\tau) : \nabla \mathbf{u}(\mathbf{x},\tau)\, d\tau \right] dt\, d^3\mathbf{x} \quad \text{(A.10)}$$

and

$$I_2 := \int\limits_G \int\limits_{t=t_0}^{t_1} [\nabla \boldsymbol{\psi}(\mathbf{x},t)] : \left[\int\limits_{-\infty}^{\infty} \boldsymbol{\Phi'}(\mathbf{x},t-\tau) : \nabla \mathbf{u}(\mathbf{x},\tau)\, d\tau \right] dt\, d^3\mathbf{x}\,. \quad \text{(A.11)}$$

The term J_2 allows us the application of Gauss' integral theorem.

$$J_2 = -\int\limits_{\partial G} \int\limits_{t=t_0}^{t_1} \left[\boldsymbol{\psi}(\mathbf{x},t) \cdot \int\limits_{\tau=-\infty}^{\infty} \boldsymbol{\Phi'}(\mathbf{x},t-\tau) : \nabla \mathbf{u}(\mathbf{x},\tau)\, d\tau \right] \cdot d\boldsymbol{\Gamma}\, dt\,. \quad \text{(A.12)}$$

Substituting $t' = t - \tau$ in the inner integral of I_2 results in

$$I_2 = \int\limits_G \int\limits_{t=t_0}^{t_1} \left[\nabla \boldsymbol{\psi}(\mathbf{x},t) : \int\limits_{t'=-\infty}^{\infty} \boldsymbol{\Phi}(\mathbf{x},t') : \nabla \mathbf{u}(\mathbf{x},t-t')\, dt' \right] dt\, d^3\mathbf{x}\,. \quad \text{(A.13)}$$

Taking into account that $\boldsymbol{\Phi'}$ is causal and that $\mathbf{u} = \mathbf{0}$ for $t < t_0$, the temporal integration domain Ω_T can be identified as (see figure A.1)

$$\begin{aligned} \Omega_T &= \{(t,t') \in \mathbb{R}^2\,;\ t_0 \le t \le t_1\,,\ 0 \le t' \le t - t_0\} \\ &= \{(t,t') \in \mathbb{R}^2\,;\ 0 \le t' \le t_1 - t_0\,,\ t' + t_0 \le t \le t_1\}\,. \end{aligned} \quad \text{(A.14)}$$

This permits to interchange the temporal integrals as follows:

$$\begin{aligned} I_2 &= \int\limits_G \int\limits_{t'=0}^{t_1-t_0} \left[\int\limits_{t=t_0+t'}^{t_1} \nabla \boldsymbol{\psi}(\mathbf{x},t) : \boldsymbol{\Phi'}(\mathbf{x},t') : \nabla \mathbf{u}(\mathbf{x},t-t')\, dt \right] dt'\, d^3\mathbf{x} \\ &= \int\limits_G \int\limits_{t=0}^{t_1-t_0} \boldsymbol{\Phi'}(\mathbf{x},t) :: \left[\int\limits_{\tau=t_0+t}^{t_1} \nabla \boldsymbol{\psi}(\mathbf{x},\tau) \otimes \nabla \mathbf{u}(\mathbf{x},\tau - t)\, d\tau \right] dt\, d^3\mathbf{x}\,. \end{aligned} \quad \text{(A.15)}$$

In order to obtain the last identity, I simply set $t' \to t$ and $t \to \tau$. Now, I_2 may be identified with the bilinear form $\langle \boldsymbol{\Phi'}, \partial_\Phi \mathbf{L}_2^*(\boldsymbol{\psi}) \rangle_{P_2}$, i. e.,

$$I_2 = \langle \boldsymbol{\Phi'}, \partial_\Phi \mathbf{L}_2^*(\boldsymbol{\psi}) \rangle_{P_2}\,. \quad \text{(A.16)}$$

d) Assembling the terms

Finally, we arrive at

$$\begin{aligned}&\langle \boldsymbol{\psi}, \partial_u \mathbf{L}_2(\mathbf{v}_2)\rangle_U + \langle \boldsymbol{\psi}, \partial_\Phi \mathbf{L}_2(\boldsymbol{\Phi}')\rangle_U \\ =&\langle \mathbf{v}_2, \partial_u \mathbf{L}_2^*(\boldsymbol{\psi})\rangle_U + \langle \boldsymbol{\Phi}', \partial_\Phi^*(\boldsymbol{\psi})\rangle_{P_2} + J_1 + J_2\,.\end{aligned} \tag{A.17}$$

Inserting the subsidiary conditions for $\mathbf{v}_2$, we find

$$\begin{aligned}&J_1 + J_2 \\ &= -\int_{\Gamma_1}\int_{t=t_0}^{t_1} \boldsymbol{\psi}(\mathbf{x},t)\cdot \int_{\tau=-\infty}^{\infty} [\boldsymbol{\Phi}'(\mathbf{x},t-\tau) : \nabla \mathbf{u}(\mathbf{x},\tau) \\ &\qquad -\boldsymbol{\Phi}(\mathbf{x},t-\tau) : \nabla \mathbf{v}_2(\mathbf{x},\tau)]\cdot d\boldsymbol{\Gamma}\, dt \\ &+\int_{\Gamma_2}\int_{t=t_0}^{t_1} \left[\mathbf{v}_2(\mathbf{x},t)\cdot \int_{\tau=-\infty}^{\infty} \boldsymbol{\Phi}(\mathbf{x},\tau-t) : \nabla\boldsymbol{\psi}(\mathbf{x},\tau)\, d\tau\right]\cdot d\boldsymbol{\Gamma}\, dt \\ &-\int_{G}\int_{t=t_0}^{t_1}\int_{\tau=t_1}^{\infty} [\nabla \mathbf{v}_2(\mathbf{x},t)] : [\boldsymbol{\Phi}(\mathbf{x},\tau-t) : \nabla\boldsymbol{\psi}(\mathbf{x},\tau)]\, d^3\mathbf{x}\, d\tau\, dt\,.\end{aligned} \tag{A.18}$$

Obviously, we can force $J_1 + J_2$ to zero, if we require that the conditions

$$\boldsymbol{\psi}(\mathbf{x},t)|_{t>t_1} = \mathbf{0}\,, \qquad \text{(quiescent future)} \tag{A.19}$$

$$\boldsymbol{\psi}(\mathbf{x},t)|_{\mathbf{x}\in\Gamma_1} = \mathbf{0}\,, \qquad \text{(Cauchy condition)} \tag{A.20}$$

$$\mathbf{n}\cdot\left[\int_{-\infty}^{\infty} \boldsymbol{\Phi}(\mathbf{x},\tau-t) : \nabla\boldsymbol{\psi}(\mathbf{x},\tau)\, d\tau\right]\Bigg|_{\mathbf{x}\in\Gamma_2} = \mathbf{0}\,, \qquad \text{(Neumann condition)} \tag{A.21}$$

hold. This must be complemented by the already known terminal conditions

$$\boldsymbol{\psi}(\mathbf{x},t)|_{t=t_1} = \mathbf{0} \quad \text{and} \quad \frac{\partial}{\partial t}\boldsymbol{\psi}(\mathbf{x},t)\Big|_{t=t_1} = \mathbf{0}\,. \tag{A.22}$$

Collecting all terms, we arrive at

$$\begin{aligned}&\langle \boldsymbol{\psi}, \partial_\rho \mathbf{L}_1(\rho')\rangle_U + \langle \boldsymbol{\psi}, \partial_u \mathbf{L}_1(\mathbf{v}_2)\rangle_U + \langle \boldsymbol{\psi}, \partial_u \mathbf{L}_2(\mathbf{v}_2)\rangle_U + \langle \boldsymbol{\psi}, \partial_\Phi \mathbf{L}_2(\boldsymbol{\Phi}')\rangle_U \\ =&\langle \rho', \partial_\rho \mathbf{L}_1^*(\boldsymbol{\psi})\rangle_{P_1} + \langle \mathbf{v}_2, \partial_u \mathbf{L}_1^*(\boldsymbol{\psi})\rangle_U + \langle \mathbf{v}_2, \partial_u \mathbf{L}_2^*(\boldsymbol{\psi})\rangle_U + \langle \boldsymbol{\Phi}', \partial_\Phi^* \mathbf{L}_2(\boldsymbol{\psi})\rangle_{P_2} \\ =&\int_G\int_{t=t_0}^{t_1} \rho'(\mathbf{x})\boldsymbol{\psi}(\mathbf{x},t)\cdot\frac{\partial^2}{\partial t^2}\mathbf{u}(\mathbf{x},t)\, dt\, d^3\mathbf{x} + \int_G\int_{t=t_0}^{t_1} \rho(\mathbf{x})\mathbf{v}_2(\mathbf{x},t)\cdot\frac{\partial^2}{\partial t^2}\boldsymbol{\psi}(\mathbf{x},t)\, dt\, d^3\mathbf{x} \\ &-\int_G\int_{t=t_0}^{t_1} \mathbf{v}_2(\mathbf{x},t)\cdot\left[\nabla\cdot\int_{\tau=-\infty}^{\infty} \boldsymbol{\Phi}(\mathbf{x},\tau-t) : \nabla\boldsymbol{\psi}(\mathbf{x},\tau)\, d\tau\right] d^3\mathbf{x}\, dt \\ &\int_G\int_{t=0}^{t_1-t_0} \boldsymbol{\Phi}'(\mathbf{x},t) :: \left[\int_{\tau=t_0+t}^{t_1} \nabla\boldsymbol{\psi}(\mathbf{x},\tau)\otimes\nabla\mathbf{u}(\mathbf{x},\tau-t)\, d\tau\right] dt\, d^3\mathbf{x}\,.\end{aligned} \tag{A.23}$$

This proves the relations given in section 6.