Hessian kernels of seismic data functionals based upon adjoint techniques

Andreas Fichtner and Jeannot Trampert

Department of Earth Sciences, Utrecht University, Utrecht, the Netherlands. E-mail: fichtner@geo.uu.nl

Accepted 2011 January 19. Received 2011 January 12; in original form 2010 November 19

SUMMARY

We present an extension of the adjoint method that allows us to compute the second derivatives of seismic data functionals with respect to Earth model parameters. This work is intended to serve as a technical prelude to the implementation of Newton-like optimization schemes and the development of quantitative resolution analyses in time-domain full seismic waveform inversion.

The Hessian operator H applied to a model perturbation $\delta \mathbf{m}$ can be expressed in terms of four different wavefields. The forward field \mathbf{u} excited by the regular source, the adjoint field \mathbf{u}^{\dagger} excited by the adjoint source located at the receiver position, the scattered forward field $\delta \mathbf{u}$ and the scattered adjoint field $\delta \mathbf{u}^{\dagger}$. The formalism naturally leads to the notion of Hessian kernels, which are the volumetric densities of $H\delta \mathbf{m}$. The Hessian kernels appear as the superposition of (1) a first-order influence zone that represents the approximate Hessian, and (2) second-order influence zones that represent second-order scattering.

To aid in the development of physical intuition we provide several examples of Hessian kernels for finite-frequency traveltime measurements on both surface and body waves. As expected, second-order scattering is efficient only when at least one of the model perturbations is located within the first Fresnel zone of the Fréchet kernel. Second-order effects from density heterogeneities are generally negligible in transmission tomography, provided that the Earth model is parameterized in terms of density and seismic wave speeds.

With a realistic full waveform inversion for European upper-mantle structure, we demonstrate that significant differences can exist between the approximate Hessian and the full Hessian—despite the near-optimality of the tomographic model. These differences are largest for the off-diagonal elements, meaning that the approximate Hessian can lead to erroneous inferences concerning parameter trade-offs. The full Hessian, in contrast, allows us to correctly account for the effect of non-linearity on model resolution.

Key words: Inverse theory; Seismic tomography; Computational seismology; Theoretical seismology; Wave scattering and diffraction; Wave propagation.

1 INTRODUCTION

1.1 Full waveform inversion

Full waveform inversion is a tomographic technique that is based on the numerical simulation of seismic wave propagation. The purely numerical solutions of the wave equation for heterogeneous Earth models provide complete and accurate synthetic seismograms that can be exploited for high-resolution imaging in complex media.

While already initiated in the late 1970s and 1980s (e.g. Bamberger *et al.* 1977, 1982; Tarantola 1984; Gauthier *et al.* 1986; Tarantola 1988), full waveform inversion has only recently gained popularity, mainly for two reasons: rapid advances in high-performance computing and our need to reveal the structure of the Earth with increasing detail. Applications of full waveform inversion to real data have now been reported in widely varying contexts, ranging from exploration and engineering problems (e.g. Igel *et al.* 1996; Operto *et al.* 2004; Gao *et al.* 2006; Smithyman *et al.* 2009) to the imaging of regional- and continental-scale structure (e.g. Dessa *et al.* 2004; Bleibinhaus *et al.* 2007, 2009; Chen *et al.* 2007; Tape *et al.* 2009, 2010; Fichtner *et al.* 2009, 2010). Despite the success of full waveform inversion, numerous challenges remain. These include the design of more efficient optimization schemes and the rigorous quantification of model resolution and uncertainties.

GJI Seismology

1.2 Efficiency of optimization schemes and the Newton method

The efficiency of iterative schemes used to minimize the misfit \mathcal{X} between observed seismograms $\mathbf{u}_0(t)$ and synthetic seismograms $\mathbf{u}(t)$ improved continuously over the past three decades. Simplistic steepest-descent methods employed in the early-development stage of full waveform inversion (e.g. Bamberger *et al.* 1982; Gauthier *et al.* 1986) have been replaced by pre-conditioned conjugate-gradient algorithms (e.g. Mora 1987, 1988; Tape *et al.* 2007; Fichtner *et al.* 2009) and Newton-like methods (including Gauss–Newton and Levenberg–Marquardt) that use the approximate Hessian (e.g. Pratt *et al.* 1998; Epanomeritakis *et al.* 2008; Brossier *et al.* 2009). The ultimate step to be taken, is the implementation of the full Newton method where the Earth model in iteration *k*, denoted by \mathbf{m}_{k} , is updated according to

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{h}_k,\tag{1}$$

where the descent direction \mathbf{h}_k is the solution of the Newton equation

$$\mathbf{H}(\mathbf{m}_k)\,\mathbf{h}_k = -\nabla_m \,\mathcal{X}(\mathbf{m}_k). \tag{2}$$

The symbols **H** and $\nabla_m \mathcal{X}$ signify the Hessian and the Fréchet derivative of the misfit functional \mathcal{X} with respect to the Earth model. Newton's method converges quadratically provided that the initial model \mathbf{m}_0 is sufficiently close to the global optimum. It therefore has the potential to reduce the number of iterations needed to reach acceptable solutions. While common practice in non-linear optimization (e.g. Hinze *et al.* 2009), the feasibility of the Newton method in full waveform inversion has so far only been explored in 1-D synthetic studies (Santosa & Symes 1988; Pratt *et al.* 1998).

1.3 Quantification of resolution

While full waveform inversion is a promising tool, resolution estimates are generally deficient. They are mostly based on synthetic inversions for specific input structures and on the visual inspection of the tomographic images. Synthetic inversions are known to be potentially misleading even in linearized tomographies (Lévêque *et al.* 1993). Visual inspection is equally inadequate because of the very efficient psychologic trap to mistake the great detail seen in the models as an indicator of comparatively high resolution. In fact, small-scale features may easily appear in full waveform inversion because it hardly requires any explicit regularization—in contrast to classical linearized tomography that is based on the solution of large ill-conditioned linear systems.

Early attempts to analyse—and in fact define—resolution were founded on the equivalence of diffraction tomography and the first iteration of a full waveform inversion (e.g. Devaney 1984; Wu & Toksöz 1987; Mora 1989). However, this equivalence holds only in the impractical case where the misfit \mathcal{X} is equal to the L_2 waveform difference $\int (\mathbf{u} - \mathbf{u}_0)^2 dt$ (Fichtner *et al.* 2008). Furthermore, the analysis of diffraction tomography is feasible only in homogeneous or layered acoustic media.

Despite being crucial for the interpretation of the tomographic images, methods for the quantification of resolution in realistic applications of full waveform inversion do not exist so far. This absence of a quantitative means to assess the capabilities of full waveform inversion is the source of much scepticism as to whether it is really worth the effort.

Part of the problem is the inherently non-linear relation between Earth structure and seismic waveforms that leads to misfit functionals with multiple local minima. At least in the vicinity of the global optimum $\tilde{\mathbf{m}}$, characterized by $\nabla_m \mathcal{X}(\tilde{\mathbf{m}}) = \mathbf{0}$, the Hessian matrix \mathbf{H} can be used to infer the resolution of and the trade-offs between model parameters. Approximating \mathcal{X} quadratically around $\tilde{\mathbf{m}}$ yields

$$\mathcal{X}(\mathbf{m}) \approx \mathcal{X}(\widetilde{\mathbf{m}}) + \frac{1}{2} (\mathbf{m} - \widetilde{\mathbf{m}})^T \mathbf{H}(\widetilde{\mathbf{m}}) (\mathbf{m} - \widetilde{\mathbf{m}}).$$
 (3)

Furthermore accepting Gaussian statistics, allows us to define a probability density

 $\sigma(\mathbf{m}) = \operatorname{const} e^{-\mathcal{X}(\mathbf{m})} = \operatorname{const} e^{-\frac{1}{2} (\mathbf{m} - \widetilde{\mathbf{m}})^T \operatorname{H}(\widetilde{\mathbf{m}}) (\mathbf{m} - \widetilde{\mathbf{m}})}.$

Eq. (4) reveals that $\mathbf{H}(\widetilde{\mathbf{m}})$ is the inverse posterior covariance matrix in the vicinity of $\widetilde{\mathbf{m}}$ (e.g. Tarantola 2005). It follows that $\mathbf{H}(\widetilde{\mathbf{m}})$ contains all the information on resolution and trade-offs, provided that we are sufficiently close to the global optimum.

1.4 Objectives and potentials

The key role played by the Hessian in optimization and resolution analysis motivates the development of methods that allow us to compute the second derivatives of \mathcal{X} with optimal efficiency. When a discretized version of the governing equations can be solved explicitly—as in 2-D frequency-domain modelling (Pratt *et al.* 1998)—the product of **H** and an arbitrary model vector can be computed conveniently via an extension of the discrete adjoint method. However, in large-scale 3-D applications where the forward problem is solved iteratively in the time domain, the matrix formalism of the discrete adjoint method becomes impractical.

Our prime objective is therefore to generalize the continuous adjoint method to the computation of the Hessian operator H applied to an arbitrary model perturbation $\delta \mathbf{m}$. This is intended to serve as a technical preparatory step for future developments, and as a means to gain the intuition and experience needed for the meaningful solution of any inverse problem.

The potential of the method developed in the following sections goes far beyond Newton's method and the quantification of uncertainties. It may as well be used for extremal bounds analysis (Meju & Sakkas 2007; Meju 2009; Fichtner 2010), for the determination of optimal step lengths in gradient-based optimization, as a tool to assess second-order effects on seismic data functionals including finite-frequency

(4)

(7)

(9)

traveltimes (see Section 4), to study wave propagation through complex media (e.g. Baig & Dahlen 2004), or to design locally independent model parameters as it is classically done in Partitioned Waveform Inversion (Nolet 1990, 2008).

1.5 Which Hessian?

The term Hessian is frequently used as a synonym for the approximate Hessian, especially in the context of linearized tomography. For the least-squares misfit functional

$$\mathcal{X} = \frac{1}{2} \left(\mathbf{d} - \mathbf{d}_0 \right)^T (\mathbf{d} - \mathbf{d}_0), \tag{5}$$

with synthetic data \mathbf{d} and observed data \mathbf{d}_0 , the Hessian matrix \mathbf{H} is given by

$$\mathbf{H} = \mathbf{G}^T \mathbf{G} + (\mathbf{d} - \mathbf{d}_0)^T \, \nabla_m \mathbf{G},\tag{6}$$

where **G** is the Jacobian matrix of **d**, that is, $\mathbf{G} = \nabla_m \mathbf{d}$. The full Hessian **H** reduces to the approximate Hessian

$$\widetilde{\mathbf{H}} = \mathbf{G}^T \mathbf{G}$$

either when the relation between **d** and **m** is linear ($\mathbf{d} \approx \mathbf{Gm}$) or in the hypothetical case of zero misfit ($\mathbf{d} = \mathbf{d}_0$). It can, however, be shown in both 1-D synthetic inversions (Santosa & Symes 1988) and large-scale 3-D waveform tomographies (Section 5) that $\widetilde{\mathbf{H}}$ can differ substantially from **H**.

The analysis presented in the following paragraphs goes beyond the linear approximation and the assumption of small misfits. It will allow us to compute the exact second derivatives of any data or misfit functional, thus fully accounting for the generally non-linear relation between Earth models and seismic observables. The distinction between linearization and inherent non-linearity will become most apparent in Section 4 where we consider the second derivatives of finite-frequency traveltimes that are commonly considered to be nearly linearly related to Earth structure.

1.6 Outline

This paper is organized as follows. To set the stage for the computation of second derivatives, and to introduce basic concepts and notations, we start with a condensed review of the standard continuous adjoint method that allows us to express the Fréchet derivative

$$\nabla_m \mathcal{X} \,\delta \mathbf{m} = \lim_{\nu \to 0} \frac{1}{\nu} [\mathcal{X}(\mathbf{m} + \nu \,\delta \mathbf{m}) - \mathcal{X}(\mathbf{m})] \tag{8}$$

in terms of the regular wavefield **u** and the adjoint wavefield \mathbf{u}^{\dagger} (Section 2.1). The extension of the adjoint method to the computation of the Hessian operator *H* applied to a continuous model perturbation $\delta \mathbf{m}(\mathbf{x})$,

$$H\,\delta\mathbf{m}=\nabla_m\nabla_m\mathcal{X}\,\delta\mathbf{m}\,,$$

is then the topic of Section 2.2. Our approach naturally leads to the concept of Hessian kernels, defined as the volumetric densities of $H\delta \mathbf{m}$. The transition to the discrete world is made by projecting the Hessian operator H onto the basis functions of the model space to obtain the components of the Hessian matrix \mathbf{H} (Section 2.2.2). Up to that point, our development will be deliberately general. This ensures that the formalism can be applied to any PDE-governed physical system in general and to different types of wave equations in particular. Section 3 is concerned with the specific application of the previously derived equations to the 3-D viscoelastic wave equation. We provide explicit formulas for Hessian kernels and we elaborate on their interpretation in terms of second-order scattering from within secondary influence zones. We furthermore highlight relations between Hessian and Fréchet kernels. To illustrate the concept, we provide a series of examples where we computed Hessian kernels for finite-frequency traveltime measurements on both surface and body waves (Section 4). In a realistic full waveform tomography for European upper-mantle structure we demonstrate that the approximate Hessian differs significantly from the full Hessian, even in the vicinity of the optimal model (Section 5). Finally, Appendix A is intended to reveal close relationships between the discrete and continuous adjoint methods for the computation of second derivatives.

2 THE CONTINUOUS ADJOINT METHOD FOR FIRST AND SECOND DERIVATIVES

2.1 First derivatives

Variants of the continuous adjoint method applied to seismic wave propagation have been described by several authors (e.g. Tarantola 1988; Tromp *et al.* 2005; Plessix 2006; Liu & Tromp 2006; Fichtner *et al.* 2006; Chen 2011). In the interest of generality and compact notation, we derive here a formulation that is independent of the type of wave equation used. For this we consider a wavefield **u** that depends on the position vector $\mathbf{x} \in G \subset \mathbb{R}^3$, time $t \in T = [t_0, t_1]$ and on model parameters $\mathbf{m} \in \mathfrak{M}$

$\mathbf{u} = \mathbf{u}(\mathbf{m}; \mathbf{x}, t).$

(10)

The model space \mathfrak{M} contains all admissible parameters $\mathbf{m}(\mathbf{x}) = [\mathbf{m}^{(1)}(\mathbf{x}), \mathbf{m}^{(2)}(\mathbf{x}), \ldots]$. The components $\mathbf{m}^{(\alpha)}(\mathbf{x})$ may, for instance, represent the distributions of density ρ , the *P* velocity v_P or the *S* velocity v_S , that is, $\mathbf{m}(\mathbf{x}) = [\rho(\mathbf{x}), v_P(\mathbf{x}), v_S(\mathbf{x}), \ldots]$. The semicolon in eq. (10) indicates

that \mathbf{u} evolves in space and in time, whereas the model parameters are assumed to be fixed for a given realization of \mathbf{u} . The wavefield \mathbf{u} is linked via the wave equation, symbolically written as

$$\mathbf{L}(\mathbf{u},\mathbf{m}) = \mathbf{f}\,,\tag{11}$$

to external sources **f** and the model parameters **m**. It is commonly not **u** itself, but a scalar data functional $\mathcal{X}(\mathbf{m}) = \mathcal{X}[\mathbf{u}(\mathbf{m})]$ that we are interested in. The data functional can either play the role of a secondary observable or of a misfit functional that quantifies the discrepancy between observed and calculated waveforms. Without loss of generality we can write \mathcal{X} in the form of an integral over time and space, that is,

$$\mathcal{X}(\mathbf{m}) = \int_{T} \int_{G} \chi[\mathbf{u}(\mathbf{m}; \mathbf{x}, t)] dt d^{3}\mathbf{x} = \langle \chi(\mathbf{m}) \rangle, \qquad (12)$$

where we introduced (.) as a short notation for the time-space integral $\int_T \int_G (.) dt d^3 \mathbf{x}$. The derivative of $\mathcal{X}(\mathbf{m})$ with respect to \mathbf{m} in a direction $\delta \mathbf{m}$ follows from the chain rule

$$\nabla_m \mathcal{X} \,\delta \mathbf{m} = \nabla_u \mathcal{X} \,\delta \mathbf{u} = \langle \nabla_u \chi \,\delta \mathbf{u} \rangle, \tag{13}$$

where

 $\delta \mathbf{u} = \nabla_m \mathbf{u} \, \delta \mathbf{m}$

denotes the derivative of **u** with respect to **m** in the direction $\delta \mathbf{m}$. The practical difficulty of eq. (13) lies in the appearance of $\delta \mathbf{u}$ which is often hard to evaluate numerically. For a first-order finite-difference approximation of $\nabla_m \mathcal{X}$ one needs to determine $\mathbf{u}(\mathbf{m} + \epsilon \delta \mathbf{m})$ for each possible direction $\delta \mathbf{m}$. This, however, becomes infeasible in the case of numerically expensive forward problems and large model spaces. Consequently, we may not be able to compute $\nabla_m \mathcal{X}$ unless we manage to eliminate $\delta \mathbf{u}$ from eq. (13). For this purpose, we differentiate the governing eqs (11) with respect to **m**. Again invoking the chain rule for differentiation gives

$$\nabla_m \mathbf{L} \,\delta \mathbf{m} + \nabla_u \mathbf{L} \,\delta \mathbf{u} = \mathbf{0} \,. \tag{15}$$

The right-hand side of eq. (15) vanishes because the external sources **f** do not depend on the model parameters **m**. We now multiply eq. (15) by an arbitrary test function \mathbf{u}^{\dagger} and then apply the integral $\langle . \rangle$

$$\langle \mathbf{u}^{\dagger} \nabla_m \mathbf{L} \, \delta \mathbf{m} \rangle + \langle \mathbf{u}^{\dagger} \nabla_u \mathbf{L} \, \delta \mathbf{u} \rangle = 0 \,. \tag{16}$$

Adding eqs (13) and (16) gives

 $\nabla_m \mathcal{X} \,\delta\mathbf{m} = \langle \nabla_u \chi \,\delta\mathbf{u} \rangle + \left\langle \mathbf{u}^{\dagger} \nabla_u \mathbf{L} \,\delta\mathbf{u} \right\rangle + \left\langle \mathbf{u}^{\dagger} \nabla_m \mathbf{L} \,\delta\mathbf{m} \right\rangle \,. \tag{17}$

We can rewrite eq. (17) using the adjoint operators $\nabla_u \chi^{\dagger}$ and $\nabla_u L^{\dagger}$ which are defined by

$$\langle \nabla_{\boldsymbol{u}} \boldsymbol{\chi} \, \delta \boldsymbol{u} \rangle = \left\langle \delta \boldsymbol{u} \nabla_{\boldsymbol{u}} \boldsymbol{\chi}^{\dagger} \right\rangle \tag{18}$$

and

$$\left\langle \mathbf{u}^{\dagger} \nabla_{\boldsymbol{u}} \mathbf{L} \, \delta \mathbf{u} \right\rangle = \left\langle \delta \mathbf{u} \nabla_{\boldsymbol{u}} \mathbf{L}^{\dagger} \, \mathbf{u}^{\dagger} \right\rangle, \tag{19}$$

for any $\delta \mathbf{u}$ and \mathbf{u}^{\dagger} . We then obtain

$$\nabla_m \mathcal{X} \,\delta\mathbf{m} = \left\langle \delta\mathbf{u} \left(\nabla_u \chi^\dagger + \nabla_u \mathbf{L}^\dagger \mathbf{u}^\dagger \right) \right\rangle + \left\langle \mathbf{u}^\dagger \nabla_m \mathbf{L} \,\delta\mathbf{m} \right\rangle \,. \tag{20}$$

We may now eliminate $\delta \mathbf{u}$ from eq. (20) if we can determine a field \mathbf{u}^{\dagger} to satisfy

$$\nabla_{u} \mathbf{L}^{\dagger} \mathbf{u}^{\dagger} = -\nabla_{u} \chi^{\dagger} \,. \tag{21}$$

Eq. (21) is referred to as the adjoint equation of (11), and \mathbf{u}^{\dagger} and $-\nabla_{u}\chi^{\dagger}$ are the adjoint field and the adjoint source, respectively. When the solution \mathbf{u}^{\dagger} of the adjoint equation is found, then the derivative of the objective functional reduces to

$$\nabla_m \mathcal{X} \,\delta \mathbf{m} = \left\langle \mathbf{u}^{\dagger} \nabla_m \mathbf{L} \,\delta \mathbf{m} \right\rangle \,. \tag{22}$$

By construction, $\nabla_m \mathcal{X} \, \delta \mathbf{m}$ can now be computed for any differentiation direction $\delta \mathbf{m}$ without the explicit knowledge of $\delta \mathbf{u}$. This advantage comes at the price of having to find the adjoint operator $\nabla_u \mathbf{L}^{\dagger}$ and a solution to the adjoint problem (21). When the operator \mathbf{L} is linear in \mathbf{u} , it follows that $\nabla_u \mathbf{L}(\mathbf{u})\mathbf{u}^{\dagger} = \mathbf{L}^{\dagger}(\mathbf{u}^{\dagger})$, and the adjoint eq. (21) can be simplified to

$$\mathbf{L}^{\dagger}(\mathbf{u}^{\dagger}) = -\nabla_{u}\chi^{\dagger}.$$
(23)

The adjoint field \mathbf{u}^{\dagger} is then determined by the adjoint operator \mathbf{L}^{\dagger} of \mathbf{L} and the right-hand side $-\nabla_{u}\chi^{\dagger}$ that is derived from the data functional.

2.1.1 Fréchet kernels

Much of our physical intuition is based on the interpretation of sensitivity or Fréchet kernels which are defined as the volumetric densities of the Fréchet derivative $\nabla_m \mathcal{X}$

$$K_m(\mathbf{u}^{\dagger}, \mathbf{u}) = \frac{d}{dV} \nabla_m \mathcal{X} = \int_T \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\mathbf{u}, \mathbf{m}) dt \,.$$
⁽²⁴⁾

© 2011 The Authors, *GJI*, **185**, 775–798 Geophysical Journal International © 2011 RAS

(14)

In eq. (24), we introduced the dependence of the Fréchet kernel K_m on **u** and \mathbf{u}^{\dagger} explicitly in the notation. The usefulness of this seemingly unnecessary complication will become apparent the Section 2.2.1 on Hessian kernels. Using the notion of Fréchet kernels, we can recast eq. (22) as follows

$$\nabla_m \mathcal{X} \,\delta\mathbf{m} = \int_G K_m \,\delta\mathbf{m} \,d^3\mathbf{x} \,. \tag{25}$$

For the special case of an isotropic Earth model parameterized in terms of ρ , v_P and v_S , eq. (25) takes the form

$$\nabla_m \mathcal{X} \,\delta\mathbf{m} = \int_G (K_\rho \,\delta\rho + K_{v_{\rm P}} \,\delta v_{\rm P} + K_{v_{\rm S}} \,\delta v_{\rm S}) \,d^3\mathbf{x} = \int_G (\bar{K}_\rho \,\delta\ln\rho + \bar{K}_{v_{\rm P}} \,\delta\ln v_{\rm P} + \bar{K}_{v_{\rm S}} \,\delta\ln v_{\rm S}) \,d^3\mathbf{x} \,, \tag{26}$$

where $\bar{K}_{\rho} = \rho K_{\rho}$, $\bar{K}_{v_{\rm P}} = v_{\rm P} K_{v_{\rm P}}$ and $\bar{K}_{v_{\rm S}} = v_{\rm S} K_{v_{\rm S}}$ are kernels for relative perturbations. The Fréchet kernels reveal how the data functional $\mathcal{X}(\mathbf{m})$ is affected by infinitesimal changes in the model parameters. It is the study of K_m for different types of seismic waves and different data functionals that allows us to design efficient inversion schemes and to interpret the results in a physically meaningful way.

2.1.2 Translation to the discretized model space

In most applications, the model space \mathfrak{M} is discretized, meaning that the components $m^{(\alpha)}$ of the space-continuous model $\mathbf{m}(\mathbf{x}) = [m^{(1)}(\mathbf{x}),$ $m^{(2)}(\mathbf{x}), \ldots$ are expressed as a linear combination of $N < \infty$ basis functions, $b_i(\mathbf{x})$

$$m^{(\alpha)}(\mathbf{x}) = \sum_{i=1}^{N} \mu_i^{(\alpha)} b_i(\mathbf{x}), \qquad \alpha = 1, 2, \dots$$
(27)

Commonly, the basis functions are spherical harmonics, blocks, wavelets or splines. With the representation (27), the model **m** and the data functional \mathcal{X} are fully determined by the coefficients or model parameters $\mu_i^{(\alpha)}$. We are therefore interested in the partial derivatives $\partial \mathcal{X} / \partial \mu_i^{(\alpha)}$. Using the definition of the classical derivative, we find

$$\frac{\partial \mathcal{X}}{\partial \mu_i^{(\alpha)}} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\mathcal{X} \left(\dots, \mu_i^{(\alpha)} + \epsilon, \dots \right) - \mathcal{X} \left(\dots, \mu_i^{(\alpha)}, \dots \right) \right] = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\mathcal{X} \left(\dots, m^{(\alpha)} + \epsilon b_i, \dots \right) - \mathcal{X} (\dots, m^{(\alpha)}, \dots) \right]$$
$$= \nabla_{m^{(\alpha)}} \mathcal{X} b_i = \int_G K_{m^{(\alpha)}}(\mathbf{x}) b_i(\mathbf{x}) d^3 \mathbf{x}.$$
(28)

It follows from eq. (28) that the gradient in the classical sense, $\partial \mathcal{X} / \partial \mu_i^{(\alpha)}$, is given by the projection of the sensitivity kernel $K_{m^{(\alpha)}}$ onto the basis function b_i .

2.2 Second derivatives

Equipped with the machinery of the adjoint method for first derivatives, we can now compute the Hessian operator H applied to a model vector $\delta \mathbf{m}_1$, that is, $H \delta \mathbf{m}_1$. For this we first differentiate \mathcal{X} with respect to **m** in the direction $\delta \mathbf{m}_1$,

$$\nabla_m \mathcal{X} \,\delta \mathbf{m}_1 = \langle \nabla_u \chi \,\delta_1 \mathbf{u} \rangle \,, \tag{29}$$

with the derivative of the forward field

$$\delta_1 \mathbf{u} = \nabla_m \mathbf{u} \,\delta \mathbf{m}_1. \tag{30}$$

Repeating this procedure for a second direction $\delta \mathbf{m}_2$ yields

$$H(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = \nabla_m \nabla_m \mathcal{X}(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = \langle \nabla_u \nabla_u \chi(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) + \nabla_u \chi \delta_{12} \mathbf{u} \rangle,$$
(31)
with the first derivative

with the first derivative

 $\delta_2 \mathbf{u} = \nabla_m \mathbf{u} \, \delta \mathbf{m}_2$ (32)

and the second derivative of the forward wavefield

$$\delta_{12}\mathbf{u} = \nabla_m \nabla_m \mathbf{u}(\delta \mathbf{m}_1, \delta \mathbf{m}_2). \tag{33}$$

The first term on the right-hand side of eq. (31),

$$H(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = \langle \nabla_u \nabla_u \chi(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) \rangle,$$

~

is the approximate Hessian of the data functional \mathcal{X} applied to $\delta \mathbf{m}_1$ and $\delta \mathbf{m}_2$. The approximate Hessian merely involves first derivatives that we can already compute efficiently for any differentiation direction with the help of the standard adjoint method that we described in Section 2.1. The Gauss–Newton and Levenberg–Marquardt methods of non-linear optimization use \tilde{H} as a computationally less expensive substitute of the full Hessian H.

The difficulty in eq. (31) is the appearance of the second derivative δ_{12} **u** that we cannot compute efficiently for arbitrary differentiation directions $\delta \mathbf{m}_1$ and $\delta \mathbf{m}_2$. To eliminate $\delta_{12}\mathbf{u}$ from (31) we again differentiate the forward problem, $\mathbf{L}(\mathbf{u}, \mathbf{m}) = \mathbf{f}$, with respect to **m** in the first direction $\delta \mathbf{m}_1$

$$\mathbf{0} = \nabla_m \mathbf{L} \, \delta \mathbf{m}_1 + \nabla_u \mathbf{L} \, \delta_1 \mathbf{u} \, .$$

© 2011 The Authors, GJI, 185, 775-798 Geophysical Journal International © 2011 RAS (34)

780 A. Fichtner and J. Trampert

Differentiating (35) once more with respect to **m** but in the direction $\delta \mathbf{m}_2$ yields

$$\mathbf{0} = \nabla_m \nabla_m \mathbf{L}(\delta \mathbf{m}_1, \delta \mathbf{m}_2) + \nabla_u \nabla_m \mathbf{L}(\delta \mathbf{m}_1, \delta_2 \mathbf{u}) + \nabla_m \nabla_u \mathbf{L}(\delta_1 \mathbf{u}, \delta \mathbf{m}_2) + \nabla_u \nabla_u \mathbf{L}(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) + \nabla_u \mathbf{L}\delta_{12} \mathbf{u} \,. \tag{36}$$

In the next step, we multiply eq. (56) with an arbitrary test function
$$\mathbf{w}^{\dagger}$$
 and integrate over time and space

$$\mathbf{0} = (\mathbf{w}^{\mathsf{v}} \nabla_m \mathbf{v}_m \mathbf{L}(o\mathbf{m}_1, o\mathbf{m}_2)) + (\mathbf{w}^{\mathsf{v}} \nabla_u \nabla_m \mathbf{L}(o\mathbf{m}_1, o_2\mathbf{u}))$$

$$+ \langle \mathbf{w}^{\mathsf{T}} \nabla_m \nabla_u \mathbf{L}(\delta_1 \mathbf{u}, \delta \mathbf{m}_2) \rangle + \langle \mathbf{w}^{\mathsf{T}} \nabla_u \nabla_u \mathbf{L}(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) \rangle + \langle \mathbf{w}^{\mathsf{T}} \nabla_u \mathbf{L} \delta_{12} \mathbf{u} \rangle .$$
(37)

Adding (37) to (31) and rearranging terms gives

$$H(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}) = \left\langle \left(\nabla_{u} \chi + \mathbf{w}^{\dagger} \nabla_{u} \mathbf{L} \right) \delta_{12} \mathbf{u} \right\rangle + \left\langle \nabla_{u} \nabla_{u} \chi (\delta_{1} \mathbf{u}, \delta_{2} \mathbf{u}) \right\rangle + \left\langle \mathbf{w}^{\dagger} \nabla_{m} \nabla_{m} \mathbf{L} (\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}) \right\rangle + \left\langle \mathbf{w}^{\dagger} \nabla_{u} \nabla_{m} \mathbf{L} (\delta \mathbf{m}_{1}, \delta_{2} \mathbf{u}) \right\rangle + \left\langle \mathbf{w}^{\dagger} \nabla_{u} \nabla_{u} \mathbf{L} (\delta_{1} \mathbf{u}, \delta_{2} \mathbf{u}) \right\rangle + \left\langle \mathbf{w}^{\dagger} \nabla_{u} \nabla_{u} \mathbf{L} (\delta_{1} \mathbf{u}, \delta_{2} \mathbf{u}) \right\rangle .$$

$$(38)$$

We can now eliminate the second derivative of the forward field, δ_{12} **u**, from eq. (38) by imposing that the test field **w**[†] be the solution of the adjoint equation

$$\nabla_{u} \mathbf{L}^{\dagger} \mathbf{w}^{\dagger} = -\nabla_{u} \chi^{\dagger} \,. \tag{39}$$

The adjoint eq. (39) is identical to the adjoint equation for first derivatives (21), which implies $\mathbf{w}^{\dagger} = \mathbf{u}^{\dagger}$. This means that the adjoint field \mathbf{u}^{\dagger} can in practice be recycled for the computation of the Hessian. When \mathbf{u}^{\dagger} satisfies the adjoint eq. (39), then $H(\delta \mathbf{m}_1, \delta \mathbf{m}_2)$ can indeed be expressed in terms of first derivatives with respect to \mathbf{m}

$$H(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}) = \langle \nabla_{u} \nabla_{u} \chi(\delta_{1} \mathbf{u}, \delta_{2} \mathbf{u}) \rangle + \langle \mathbf{u}^{\dagger} \nabla_{m} \nabla_{m} \mathbf{L}(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}) \rangle + \langle \mathbf{u}^{\dagger} \nabla_{u} \nabla_{m} \mathbf{L}(\delta \mathbf{m}_{1}, \delta_{2} \mathbf{u}) \rangle + \langle \mathbf{u}^{\dagger} \nabla_{m} \nabla_{u} \mathbf{L}(\delta_{1} \mathbf{u}, \delta \mathbf{m}_{2}) \rangle + \langle \mathbf{u}^{\dagger} \nabla_{u} \nabla_{u} \mathbf{L}(\delta_{1} \mathbf{u}, \delta_{2} \mathbf{u}) \rangle .$$

$$(40)$$

The last term on the right-hand side of eq. (40), involving $\nabla_u \nabla_u \mathbf{L}$, is zero for linear operators, including the wave equation operator. Whether the second term, involving $\nabla_m \nabla_m \mathbf{L}$, is zero or not depends on the specific parameterization of the model space. The wave equation operator, for instance, is linear in density and the elastic parameters which leads to a zero second derivative, that is, $\nabla_m \nabla_m \mathbf{L} = \mathbf{0}$. However, when the model is parameterized in terms of density and seismic wave speeds, quadratic terms appear and $\nabla_m \nabla_m \mathbf{L}$ is generally non-zero.

To apply conjugate gradient type methods to the solution of the Newton eq. (2) we require the Hessian applied to a model perturbation, that is $H\delta \mathbf{m}_1$ and not $H(\delta \mathbf{m}_1, \delta \mathbf{m}_2)$. Unfortunately, $\delta \mathbf{m}_2$ appears implicity via $\delta_2 \mathbf{u}$ in eq. (40). The next step is therefore the isolation of $\delta \mathbf{m}_2$. For clarity, we restrict the following development to operators $\mathbf{L}(\mathbf{u})$ that are linear in \mathbf{u} , and we omit the dependence of \mathbf{L} on \mathbf{m} in the notation. This gives

$$\nabla_{\boldsymbol{u}} \nabla_{\boldsymbol{u}} \mathbf{L}(\mathbf{u})(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) = \mathbf{0} \,, \tag{41}$$

$$\nabla_{u}\nabla_{m}\mathbf{L}(\mathbf{u})(\delta\mathbf{m}_{1},\delta_{2}\mathbf{u}) = \nabla_{m}\mathbf{L}(\delta_{2}\mathbf{u})\,\delta\mathbf{m}_{1}\,,\tag{42}$$

$$\nabla_m \nabla_u \mathbf{L}(\mathbf{u})(\delta_1 \mathbf{u}, \, \delta \mathbf{m}_2) = \nabla_m \mathbf{L}(\delta_1 \mathbf{u}) \, \delta \mathbf{m}_2 \, .$$

After slight rearrangements we can now write eq. (40) in the following form

 $H(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = \langle \nabla_u \nabla_u \chi(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) \rangle + \left\langle \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\delta_2 \mathbf{u}) \, \delta \mathbf{m}_1 \right\rangle + \left\langle \mathbf{u}^{\dagger} \nabla_m \nabla_m \mathbf{L}(\mathbf{u}) (\delta \mathbf{m}_1, \delta \mathbf{m}_2) \right\rangle + \left\langle \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\delta_1 \mathbf{u}) \, \delta \mathbf{m}_2 \right\rangle. \tag{44}$

Our focus is now on the first two terms on the right-hand side of eq. (44) where the direction $\delta \mathbf{m}_2$ appears implicitly via $\delta_2 \mathbf{u} = \nabla_m \mathbf{u} \, \delta \mathbf{m}_2$. Invoking the adjoint $\nabla_m \mathbf{L}^{\dagger}$ of $\nabla_m \mathbf{L}$ we can write

$$\langle \nabla_{u} \nabla_{u} \chi(\delta_{1} \mathbf{u}, \delta_{2} \mathbf{u}) \rangle + \left\langle \mathbf{u}^{\dagger} \nabla_{m} \mathbf{L}(\delta_{2} \mathbf{u}) \, \delta \mathbf{m}_{1} \right\rangle = \left\langle \delta_{2} \mathbf{u} \left[\nabla_{u} \nabla_{u} \chi^{\dagger}(\delta_{1} \mathbf{u}) + \nabla_{m} \mathbf{L}^{\dagger}(\mathbf{u}^{\dagger}) \, \delta \mathbf{m}_{1} \right] \right\rangle. \tag{45}$$

In eq. (45) we have recognized that $\nabla_u \nabla_u \chi(\delta_1 \mathbf{u}, \delta_2 \mathbf{u})$ can be interpreted as a linear operator, $\nabla_u \nabla_u \chi(\delta_1 \mathbf{u})$, acting on $\delta_2 \mathbf{u}$, that is, $\nabla_u \nabla_u \chi(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) = \nabla_u \nabla_u \chi(\delta_1 \mathbf{u}) \delta_2 \mathbf{u}$. To further simplify eq. (45), we differentiate the adjoint eq. (39) with respect to **m** in the direction $\delta \mathbf{m}_1$, keeping in mind that \mathbf{L}^{\dagger} is linear in \mathbf{u}^{\dagger}

$$\mathbf{L}^{\dagger}\left(\delta_{1}\mathbf{u}^{\dagger}\right) = -\nabla_{u}\nabla_{u}\chi^{\dagger}(\delta_{1}\mathbf{u}) - \nabla_{m}\mathbf{L}^{\dagger}(\mathbf{u}^{\dagger})\delta\mathbf{m}_{1}.$$
(46)

Eq. (46) reveals that the derivative of the adjoint field $\delta_1 \mathbf{u}^{\dagger} = \nabla_m \mathbf{u}^{\dagger} \delta \mathbf{m}_1$ is the solution of an adjoint equation with two source terms on the right-hand side. The first source term, $-\nabla_u \nabla_u \chi^{\dagger}(\delta_1 \mathbf{u})$, is the derivative of the regular adjoint source $-\nabla_u \chi^{\dagger}$ that we already encountered in eqs (21) and (23). It accounts for the change of the adjoint source that results from the perturbation of the forward wavefield \mathbf{u} . The second source term, $-\nabla_m \mathbf{L}^{\dagger}(\mathbf{u}^{\dagger}) \delta \mathbf{m}_1$, is located at the perturbation $\delta \mathbf{m}_1$ itself, and it is responsible for the excitation of a scattered adjoint wavefield (see Figs 2 and 5 for illustrations). We will come back to the physical interpretation of $\delta_1 \mathbf{u}^{\dagger}$ and its two sources in Section 3.3.2.

With the help of eq. (46) we can simplify eq. (45)

-

$$\langle \nabla_{u} \nabla_{u} \chi(\delta_{1} \mathbf{u}, \delta_{2} \mathbf{u}) \rangle + \langle \mathbf{u}^{\dagger} \nabla_{m} \mathbf{L}(\delta_{2} \mathbf{u}) \, \delta \mathbf{m}_{1} \rangle = - \langle \delta_{2} \mathbf{u} \mathbf{L}^{\dagger}(\delta_{1} \mathbf{u}^{\dagger}) \rangle = - \langle \delta_{1} \mathbf{u}^{\dagger} \mathbf{L}(\delta_{2} \mathbf{u}) \rangle . \tag{47}$$

To replace $\mathbf{L}(\delta_2 \mathbf{u})$ in eq. (47) we note that the differentiation of the forward problem, $\mathbf{L}(\mathbf{u}, \mathbf{m}) = \mathbf{f}$, with respect to \mathbf{m} in the direction $\delta \mathbf{m}_2$ gives

$$\nabla_m \mathbf{L}(\mathbf{u}) \,\delta \mathbf{m}_2 + \mathbf{L}(\delta_2 \mathbf{u}) = \mathbf{0} \,. \tag{48}$$

Inserting (48) into (47) leaves us with the following expression

$$\langle \nabla_{\boldsymbol{u}} \nabla_{\boldsymbol{u}} \boldsymbol{\chi}(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) \rangle + \left\langle \mathbf{u}^{\dagger} \nabla_{\boldsymbol{m}} \mathbf{L}(\delta_2 \mathbf{u}) \, \delta \mathbf{m}_1 \right\rangle = \left\langle \delta_1 \mathbf{u}^{\dagger} \nabla_{\boldsymbol{m}} \mathbf{L}(\mathbf{u}) \, \delta \mathbf{m}_2 \right\rangle, \tag{49}$$

(43)

(10)

where the second differentiation direction, $\delta \mathbf{m}_2$, appears explicitly. The substitution of (49) into eq. (44) now allows us to assemble the Hessian operator applied to $\delta \mathbf{m}_1$, that is, $H\delta \mathbf{m}_1$

$$H\delta\mathbf{m}_{1} = H(\delta\mathbf{m}_{1}, \circ) = \left\langle \delta_{1}\mathbf{u}^{\dagger}\nabla_{m}\mathbf{L}(\mathbf{u})\circ\right\rangle + \left\langle \mathbf{u}^{\dagger}\nabla_{m}\mathbf{L}(\delta_{1}\mathbf{u})\circ\right\rangle + \left\langle \mathbf{u}^{\dagger}\nabla_{m}\nabla_{m}\mathbf{L}(\mathbf{u})(\delta\mathbf{m}_{1})\circ\right\rangle.$$
(50)

The symbol \circ represents a place holder for a second differentiation direction $\delta \mathbf{m}_2$.

2.2.1 Hessian kernels

Eq. (50) suggests the representation of $H(\delta \mathbf{m}_1, \delta \mathbf{m}_2)$ in terms of volumetric integral kernels that are reminiscent of the Fréchet kernels introduced in eqs (24) and (25)

$$H\left(\delta\mathbf{m}_{1}, \delta\mathbf{m}_{2}\right) = \int_{G} K_{m}^{1} \,\delta\mathbf{m}_{2} \,d^{3}\mathbf{x} = \int_{G} \left(K_{m}^{2 \to 1} + K_{m}^{1 \leftrightarrow 1} + K_{m}^{1 \to 2}\right) \,\delta\mathbf{m}_{2} \,d^{3}\mathbf{x},\tag{51}$$

with the Hessian kernels defined by

$$K_m^1 = K_m^{2 \to 1} + K_m^{1 \to 1} + K_m^{1 \to 2},$$
(52)

$$K_m^{2\to 1} = \int_T \delta_1 \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\mathbf{u}) \, dt \,, \tag{53}$$

$$K_m^{1\leftrightarrow 1} = \int_T \mathbf{u}^{\dagger} \nabla_m \nabla_m \mathbf{L}(\mathbf{u})(\delta \mathbf{m}_1) \, dt \,, \tag{54}$$

$$K_m^{1\to2} = \int_T \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\delta_1 \mathbf{u}) \, dt \,. \tag{55}$$

In this sense, we may identify the Hessian operator H applied to the model perturbation $\delta \mathbf{m}_1$ with the Hessian kernels, that is

$$H\,\delta\mathbf{m}_1 \equiv K_m^1\,. \tag{56}$$

Considering again the special case of an isotropic medium parameterized in terms of ρ , v_P and v_S , eq. (51) can be written as

$$H(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}) = \int_{G} \left(K_{\rho}^{1} \,\delta \rho_{2} + K_{\nu_{p}}^{1} \,\delta v_{P2} + K_{\nu_{s}}^{1} \,\delta v_{S2} \right) \,d^{3} \mathbf{x} = \int_{G} \left(\bar{K}_{\rho}^{1} \,\delta \ln \rho_{2} + \bar{K}_{\nu_{p}}^{1} \,\delta \ln v_{P2} + \bar{K}_{\nu_{s}}^{1} \,\delta \ln v_{S2} \right) \,d^{3} \mathbf{x} \,, \tag{57}$$

with the kernels for relative perturbations $\bar{K}_{\rho}^{1} = \rho K_{\rho}^{1}$, $\bar{K}_{\nu p}^{1} = v_{P} K_{\nu p}^{1}$ and $\bar{K}_{\nu s}^{1} = v_{S} K_{\nu s}^{1}$. Unlike Fréchet kernels, Hessian kernels depend on the model perturbation, $\delta \mathbf{m}_{1}$, either explicitly as in $K_{m}^{1 \rightarrow 1}$ or implicitly via $\delta_{1} \mathbf{u}$ and $\delta_{1} \mathbf{u}^{\dagger}$. The comparison of eqs (53) and (55) with eq. (24) reveals that the Hessian kernels $K_{m}^{2 \rightarrow 1}$ and $K_{m}^{1 \rightarrow 2}$ are closely related to the Fréchet kernel K_{m} . In fact, we find

$$K_m^{2 \to 1} = K_m \left(\delta_1 \mathbf{u}^{\dagger}, \mathbf{u} \right) , \tag{58}$$

and

$$K_m^{1\to 2} = K_m \left(\mathbf{u}^{\dagger}, \delta_1 \mathbf{u} \right) \,. \tag{59}$$

The explicit formulas derived in Section 3.2 for the Fréchet kernels with respect to specific Earth model parameters can thus be reused for the computation of the Hessian kernels $K_m^{2\rightarrow 1}$ and $K_m^{1\rightarrow 2}$. Eqs (58) and (59) are also important for the physical interpretation of the Hessian kernels.

2.2.2 Translation to the discretized model space

In a discretized model space the components $\mathbf{m}^{(\alpha)}(\mathbf{x})$ of the space-continuous model $\mathbf{m}(\mathbf{x})$ are expressed in terms of a linear combination of a finite number of basis functions, $b_i(\mathbf{x})$, as introduced in eq. (27). With the help of the Hessian operator H we can then compute the components of the Hessian matrix \mathbf{H} . The component $(\mathbf{H})_{ij}^{(\alpha\beta)}$ of \mathbf{H} is the second derivative of the data functional \mathcal{X} with respect to the discrete model parameters $\mu_i^{(\alpha)}$ and $\mu_i^{(\beta)}$, that is

$$(\mathbf{H})_{ij}^{(\alpha\beta)} = \frac{\partial^2 \mathcal{X}}{\partial \mu_i^{(\alpha)} \partial \mu_j^{(\beta)}} \,. \tag{60}$$

Making use of eq. (28) we find

$$(\mathbf{H})_{ij}^{(\alpha\beta)} = \frac{\partial}{\partial \mu_j^{(\beta)}} \left(\frac{\partial \mathcal{X}}{\partial \mu_i^{(\alpha)}} \right) = \frac{\partial}{\partial \mu_j^{(\beta)}} (\nabla_{m^{(\alpha)}} \mathcal{X} b_i) = \nabla_{m^{(\beta)}} \nabla_{m^{(\alpha)}} \mathcal{X}(b_i, b_j).$$
(61)

The component $(\mathbf{H})_{ij}^{(\alpha\beta)}$ of the Hessian matrix is therefore equal to the Hessian operator $\nabla_{m(\beta)} \nabla_{m(\alpha)} \mathcal{X}$ applied to the basis functions b_i and b_j . It follows that the Hessian kernel K_m^1 for the model perturbation $\delta \mathbf{m}_1 = b_i$ can be interpreted as a continuous representation of the *i*th row of the Hessian matrix **H**.

The derivation of the adjoint equations in the continuous space followed by the discretization through projection is characteristic for time-domain full waveform inversion. This approach differs from frequency-domain full waveform inversion where the discrete adjoint

method is applied to a semi-discrete version of the governing equations. That the continuous and discrete adjoint methods for both first and second derivatives are closely related will be shown in Appendix A.

3 APPLICATION TO THE SEISMIC WAVE EQUATION

Our developments have so far been general in the sense that the equations do not depend on the particular type of wave equation used to model seismic wave propagation. In the following paragraphs, we will be more specific and apply the adjoint formalism to the 3-D viscoelastic wave equation.

3.1 Governing equations and their adjoints

The propagation of seismic waves in the Earth can be modelled with the seismic wave equation

 $\mathbf{L}(\mathbf{u},\mathbf{m}) = \rho(\mathbf{x})\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla \sigma(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t), \quad \mathbf{x} \in G \subset \mathbb{R}^3, \quad t \in [0, t_e] \subset \mathbb{R}, \quad \mathbf{m} = (\rho, \mathbf{C}),$ (62)

that relates the displacement field **u** in the Earth $G \subset \mathbb{R}^3$ to its mass density ρ , the stress tensor σ and an external force density **f** [see Dahlen & Tromp (1998), Kennett (2001) or Aki & Richards (2002) for detailed derivations of eq. 62]. The wave operator **L** is linear in **u**. The free surface boundary condition imposes that the normal components of the stress tensor σ vanish at the surface ∂G of the Earth, that is, $\sigma \mathbf{n}|_{x\in\partial G} = \mathbf{0}$, where **n** is the unit normal on ∂G . Both the displacement field **u** and the velocity field $\dot{\mathbf{u}}$ are required to satisfy the initial condition of being equal to zero prior to t = 0: $\mathbf{u}|_{t\leq 0} = \mathbf{0}$. To obtain a complete set of equations, the stress tensor σ must be related to the displacement field **u**. For this we assume that σ depends linearly on the history of the strain tensor $\boldsymbol{\epsilon} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$

$$\boldsymbol{\sigma}(\mathbf{x},t) = \int_{-\infty}^{\infty} \dot{\mathbf{C}}(\mathbf{x},t-\tau) : \boldsymbol{\epsilon}(\mathbf{x},\tau) d\tau .$$
(63)

The symbol : in eq. (63) denotes the contraction over the two innermost indices, that is, $(\dot{\mathbf{C}} : \boldsymbol{\epsilon})_{ij} = \sum_{k,l=1}^{3} \dot{C}_{ijkl} \boldsymbol{\epsilon}_{kl}$. Eq. (63) defines a linear viscoelastic rheology. The fourth-order tensor \mathbf{C} is the elastic tensor. Since the current stress cannot depend on future strain, we require the elastic tensor to be causal: $\mathbf{C}(t)|_{t<0} = \mathbf{0}$. The symmetry of $\boldsymbol{\epsilon}$, the conservation of angular momentum and the relation of \mathbf{C} to the internal energy (Aki & Richards 2002) require that the components of \mathbf{C} satisfy a set of symmetry relations, $C_{ijkl} = C_{klij} = C_{jikl}$, that reduce the number of independent components to 21. The number of non-zero independent elastic tensor components, or elastic parameters, determines the anisotropic properties of the elastic medium (e.g. Babuška & Cara 1991).

The adjoint wave operator L^{\dagger} , that governs the propagation of the adjoint wavefield, is given by (Tarantola 1988; Tromp *et al.* 2005; Plessix 2006; Fichtner *et al.* 2006)

$$\mathbf{L}^{\dagger}(\mathbf{u}^{\dagger},\mathbf{m}) = \rho(\mathbf{x})\ddot{\mathbf{u}}^{\dagger}(\mathbf{x},t) - \nabla\sigma^{\dagger}(\mathbf{x},t), \tag{64}$$

with the adjoint stress tensor defined by

$$\boldsymbol{\sigma}^{\dagger}(\mathbf{x},t) = \int_{-\infty}^{\infty} \dot{\mathbf{C}}(\mathbf{x},\tau-t) : \nabla \mathbf{u}^{\dagger}(\mathbf{x},\tau) \, d\tau \,. \tag{65}$$

Just as the regular stress tensor σ , the adjoint stress tensor σ^{\dagger} is required to satisfy the free surface boundary condition $\sigma^{\dagger}\mathbf{n}|_{x\in\partial G} = \mathbf{0}$. Furthermore, the propagation of the adjoint wavefield is constrained by the terminal conditions $\mathbf{u}^{\dagger}|_{t\geq t_e} = \dot{\mathbf{u}}^{\dagger}|_{t\geq t_e} = \mathbf{0}$, where t_e is the time when the observation ends. In non-dissipative media the elastic wave operator \mathbf{L} is self-adjoint, meaning that $\mathbf{L} = \mathbf{L}^{\dagger}$. The obvious numerical difficulty in solving the adjoint equation is the occurrence of the terminal conditions that require that the adjoint field be zero at time $t = t_e$. In practice, this condition can only be met by solving the adjoint equation backwards in time, that is by reversing the time axis from $0 \rightarrow t_e$ to $t_e \rightarrow 0$. The terminal conditions then act as zero initial conditions, at least in the numerical simulation. Eq. (65) reveals that the adjoint stress tensor σ^{\dagger} at time t depends on future strain from t to t_e . This results in a growth of elastic energy when the wavefield propagates in the regular time direction from 0 to t_e . In reversed time, however, the elastic energy decays, so that numerical instabilities do not occur (Tarantola 1988). For efficient strategies to solve the adjoint equation and the time integral that appears in the Fréchet kernels, the reader is referred to Griewank & Walther (2000), Liu & Tromp (2006) or Fichtner *et al.* (2009).

3.2 First derivatives and Fréchet kernels

The most general expression for the derivative of a data functional $\mathcal{X}(\mathbf{m})$ in the direction $\delta \mathbf{m}$ is given by eq. (22) that we repeat here for convenience

$$\nabla_m \mathcal{X} \,\delta \mathbf{m} = \langle \mathbf{u}^{\dagger} \nabla_m \mathbf{L} \,\delta \mathbf{m} \rangle \,. \tag{66}$$

Substituting the governing eqs (62) and (63) for the general operator L yields the explicit formula

$$\nabla_m \mathcal{X} \,\delta \mathbf{m} = \int_T \int_G \mathbf{u}^{\dagger}(t) \left[\delta \rho \,\ddot{\mathbf{u}}(t) - \nabla \int_{\tau = -\infty}^{\infty} \delta \dot{\mathbf{C}}(t - \tau) : \nabla \mathbf{u}(\tau) \,d\tau \right] dt \,d^3 \mathbf{x} \,, \tag{67}$$

© 2011 The Authors, *GJI*, **185**, 775–798 Geophysical Journal International © 2011 RAS



Figure 1. Schematic illustration of the primary influence zone where the regular wavefield **u** interacts with the adjoint wavefield \mathbf{u}^{\dagger} . Numbers are used to mark the regular and adjoint wave fronts at successive points in time. As time goes on, the regular wavefield propagates away from the source while the adjoint wavefield collapses into the receiver. In numerical simulations the adjoint equations are solved backwards in time to satisfy the terminal conditions. On the reverse time axis, the adjoint field propagates away from the receiver, starting at the final observation time. The primary influence zone marks the region where a model perturbation $\delta \mathbf{m}$ generates a first-order scattered wavefield that affects the measurement at the receiver. Perturbations located outside the primary influence zone have no first-order effect on the measurement. The spatial extension of the primary influence zone is proportional to the length of the analysis time window considered in the seismograms.

with the model perturbation $\delta \mathbf{m} = (\delta \rho, \delta \mathbf{C})$. To avoid clutter, we omitted spatial dependencies in the notation. Integrating by parts provides a more symmetric and more useful version of eq. (67)

$$\nabla_m \mathcal{X} \,\delta \mathbf{m} = -\int_T \int_G \delta\rho \,\dot{\mathbf{u}}^{\dagger}(t) \dot{\mathbf{u}}(t) \,dt \,d^3 \mathbf{x} + \int_T \int_G \int_{\tau=-\infty}^{\infty} \boldsymbol{\epsilon}^{\dagger}(t) : \delta \dot{\mathbf{C}}(t-\tau) : \boldsymbol{\epsilon}(\tau) \,d\tau \,dt \,d^3 \mathbf{x} \,, \tag{68}$$

where the adjoint strain tensor ϵ^{\dagger} is defined by $\epsilon^{\dagger} = \frac{1}{2} [(\nabla \mathbf{u}^{\dagger}) + (\nabla \mathbf{u}^{\dagger})^{T}]$. The Fréchet kernels associated with (68) are

$$K_{\rho} = -\int_{T} \dot{\mathbf{u}}^{\dagger}(t) \dot{\mathbf{u}}(t) dt , \qquad (69)$$

and

$$\mathbf{K}_{C}(\tau) = \int_{T} \boldsymbol{\epsilon}^{\dagger}(t) \otimes \boldsymbol{\epsilon}(t+\tau) \, dt \,. \tag{70}$$

The symbol \otimes signifies the tensor or dyadic product. Both kernels are non-zero only within the primary influence zone where the regular and adjoint wavefields interact at some time between t = 0 and $t = t_e$. The primary influence zone, illustrated in Fig. 1, is the region where a model perturbation $\delta \mathbf{m}$ causes the regular wavefield \mathbf{u} to generate a first-order or single-scattered wave that affects the measurement at the receiver. A perturbation located outside the primary influence zone has no first-order effect on the measurement.

For most seismic phases, the primary influence zone is a roughly cigar-shaped region connecting the source and the receiver. Its precise geometry depends on many factors including the frequency content, the length of the considered time window, the type of measurement and the reference Earth model, **m**. Specific examples for common seismic phases and measurements can be found, for instance, in Friederich (1999), Zhou *et al.* (2004), Yoshizawa & Kennett (2005), Zhao *et al.* (2005), Liu & Tromp (2006, 2008), Nissen-Meyer *et al.* (2007), Sieminski *et al.* (2007a,b) or Zhou (2009).

3.2.1 Perfectly elastic and isotropic medium

Eq. (68) is of general validity. The most relevant special case, that we will use for illustration, is the perfectly elastic and isotropic medium. Perfect elasticity means that the time-dependence of the elastic tensor C and its perturbation δ C has the form of a unit-step or Heaviside function H(t)

$$\mathbf{C}(\mathbf{x},t) = \mathbf{C}(\mathbf{x}) H(t), \quad \delta \mathbf{C}(\mathbf{x},t) = \delta \mathbf{C}(\mathbf{x}) H(t).$$

© 2011 The Authors, *GJI*, **185**, 775–798 Geophysical Journal International © 2011 RAS

784 A. Fichtner and J. Trampert

Upon inserting (71) into eq. (68) we obtain a simplified expression for $\nabla_m \mathcal{X}$

$$\nabla_m \mathcal{X} \,\delta \mathbf{m} = -\int_T \int_G \delta \rho \,\dot{\mathbf{u}}^{\dagger}(t) \dot{\mathbf{u}}(t) \,dt \,d^3 \mathbf{x} + \int_T \int_G \boldsymbol{\epsilon}^{\dagger}(t) : \delta \mathbf{C} : \boldsymbol{\epsilon}(t) \,dt \,d^3 \mathbf{x} \,. \tag{72}$$
In an isotropic medium the components of \mathbf{C} are given by (Aki & Richards 2002)

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu \delta_{ik} \delta_{jl} + \mu \delta_{il} \delta_{jk} \,. \tag{73}$$

The symbols λ and μ denote the Lamé parameters. It follows that the complete derivative of \mathcal{X} is composed of three terms

$$abla_m \mathcal{X} \, \delta \mathbf{m} =
abla_
ho \mathcal{X} \, \delta
ho +
abla_\lambda \mathcal{X} \, \delta \lambda +
abla_\mu \mathcal{X} \, \delta \mu \, ,$$

with

$$\nabla_{\rho} \mathcal{X} \,\delta\rho = -\int_{T} \int_{G} \delta\rho \,\dot{\mathbf{u}}^{\dagger} \dot{\mathbf{u}} \,dt \,d^{3}\mathbf{x} \,, \tag{75a}$$

$$\nabla_{\lambda} \mathcal{X} \,\delta\lambda = \int_{T} \int_{G} \delta\lambda \,(\mathrm{tr}\,\boldsymbol{\epsilon})(\mathrm{tr}\,\boldsymbol{\epsilon}^{\dagger}) \,dt \,d^{3}\mathbf{x}\,,\tag{75b}$$

$$\nabla_{\mu} \mathcal{X} \delta \mu = 2 \int_{T} \int_{G} \delta \mu \, \boldsymbol{\epsilon} : \boldsymbol{\epsilon}^{\dagger} \, dt \, d^{3} \mathbf{x} \,. \tag{75c}$$

The symbol tr ϵ signifies the trace of the strain tensor ϵ . The associated Fréchet kernels are

$$K^0_{\rho} = -\int_T \dot{\mathbf{u}}^{\dagger} \dot{\mathbf{u}} \, dt \,, \tag{76a}$$

$$K_{\lambda}^{0} = \int_{T} (\operatorname{tr} \boldsymbol{\epsilon})(\operatorname{tr} \boldsymbol{\epsilon}^{\dagger}) dt , \qquad (76b)$$

$$K^0_{\mu} = 2 \int_T \boldsymbol{\epsilon} : \boldsymbol{\epsilon}^{\dagger} dt .$$
(76c)

The superscript ⁰ symbolizes that the Fréchet kernels correspond to the fundamental parameterization $\mathbf{m} = (\rho, \lambda, \mu)$. Based on eqs (75) and (76) we can deduce the Fréchet kernels for a perfectly elastic and isotropic medium that is parameterized in terms of density ρ , the *S* wave speed $v_{\rm S} = \sqrt{\mu/\rho}$ and the *P* wave speed $v_{\rm P} = \sqrt{(\lambda + 2\mu)/\rho}$

$$K_{\rho} = K_{\rho}^{0} + \left(v_{\rm P}^{2} - 2v_{\rm S}^{2}\right) K_{\lambda}^{0} + v_{\rm S}^{2} K_{\mu}^{0}, \tag{77a}$$

$$K_{v_{\rm S}} = 2\rho v_{\rm S} \, K^0_\mu - 4\rho v_{\rm S} \, K^0_\lambda \,, \tag{77b}$$

$$K_{\nu \mathrm{p}} = 2\rho v_{\mathrm{P}} K_{\lambda}^{0}. \tag{77c}$$

3.3 Second derivatives and Hessian kernels

As we have seen already in Section 2.2.1, the second derivative $H(\delta \mathbf{m}_1, \delta \mathbf{m}_2)$ can be conveniently expressed in terms of three Hessian kernels, $K_m^{2\rightarrow 1}$, $K_m^{1\rightarrow 2}$ and $K_m^{1\rightarrow 1}$. It follows from eqs (58) and (59) that we can reuse the formulas for Fréchet kernels from Section 3.2 to compute the Hessian kernels $K_m^{2\rightarrow 1}$ and $K_m^{1\rightarrow 2}$ for specific Earth model parameters. The only difference is the involvement of the scattered adjoint field $\delta_1 \mathbf{u}^{\dagger}$ and the scattered forward field $\delta_1 \mathbf{u}$ in the kernel calculations.

Some additional work is required for the kernel $K_m^{1\leftrightarrow 1}$: As a preparatory step towards explicit expressions for special rheologies, we substitute the governing eqs (62) and (63) into the definition of $K_m^{1\leftrightarrow 1}$ (eq. 54)

$$\int_{G} K_{m}^{1\leftrightarrow1} \delta \mathbf{m}_{2} d^{3} \mathbf{x} = \langle \mathbf{u}^{\dagger} \nabla_{m} \nabla_{m} \mathbf{L}(\delta \mathbf{m}_{1}) \delta \mathbf{m}_{2} \rangle$$

$$= \int_{T} \int_{G} \nabla_{m} \nabla_{m} \rho \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2} \right) \mathbf{u}^{\dagger}(t) \ddot{\mathbf{u}}(t) dt d^{3} \mathbf{x} - \int_{T} \int_{G} \mathbf{u}^{\dagger}(t) \left[\nabla \int_{\tau=-\infty}^{\infty} \nabla_{m} \nabla_{m} \dot{\mathbf{C}}(t-\tau) (\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}) : \nabla \mathbf{u}(\tau) \right] d\tau dt d^{3} \mathbf{x} .$$
(78)

In eq. (78) we did not specify a particular parameterization. This is accounted for by the notations $\nabla_m \nabla_m \rho$ and $\nabla_m \nabla_m \dot{\mathbf{C}}$ that allow ρ and \mathbf{C} to depend on model parameters \mathbf{m} other than density and the elastic coefficients themselves. Integrating eq. (78) by parts and writing the resulting expression in terms of the regular and adjoint strain fields, $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}^{\dagger}$, gives

$$\int_{G} K_{m}^{1\leftrightarrow1} \delta \mathbf{m}_{2} d^{3} \mathbf{x}$$

$$= -\int_{T} \int_{G} \nabla_{m} \nabla_{m} \rho \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}\right) \dot{\mathbf{u}}_{1}^{\dagger}(t) \dot{\mathbf{u}}(t) dt d^{3} \mathbf{x} + \int_{T} \int_{G} \int_{\tau=t_{0}}^{t} \boldsymbol{\epsilon}^{\dagger}(t) : \left[\nabla_{m} \nabla_{m} \dot{\mathbf{C}}(t-\tau) \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2}\right)\right] : \boldsymbol{\epsilon}(\tau) d\tau dt d^{3} \mathbf{x}.$$
(79)

Eq. (79) is, just as eq. (68) for the first derivative $\nabla_m \mathcal{X} \delta \mathbf{m}$, of general validity. Special cases can be derived by further specifications of the rheology, that is, the elastic tensor components and their time dependence. Again, we consider the perfectly elastic and isotropic case as an example.

(74)

3.3.1 Perfectly elastic and isotropic medium

First, we substitute the perfectly elastic and isotropic rheology, as described by eqs (71) and (73), into the general expression (79)

$$\int_{G} K_{m}^{1 \leftrightarrow 1} \delta \mathbf{m}_{2} d^{3} \mathbf{x} = -\int_{T} \int_{G} \nabla_{m} \nabla_{m} \rho \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2} \right) \dot{\mathbf{u}}^{\dagger} \dot{\mathbf{u}} dt d^{3} \mathbf{x} + \int_{T} \int_{G} \nabla_{m} \nabla_{m} \lambda \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2} \right) (\mathrm{tr} \, \boldsymbol{\epsilon}^{\dagger}) (\mathrm{tr} \boldsymbol{\epsilon}) dt d^{3} \mathbf{x} + \int_{T} \int_{G} \nabla_{m} \nabla_{m} \mu \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2} \right) \boldsymbol{\epsilon}^{\dagger} : \boldsymbol{\epsilon} dt d^{3} \mathbf{x}.$$
(80)

When the isotropic medium is parameterized in terms of ρ , λ and μ or, alternatively, ρ , κ and μ , we find

$$K_m^{1\leftrightarrow 1} = 0, (81)$$

meaning that the Hessian kernel $K_{\rm m}^{1\leftrightarrow 1}$ is identically zero. This result depends critically on the choice of the free parameters. Changing, for instance, the set of free parameters to ρ , $v_{i\rm P}$ and $v_{i\rm S}$, results in

$$\nabla_{m} \nabla_{m} \rho \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2} \right) = 0,$$

$$\nabla_{m} \nabla_{m} \mathbf{C} \left(\delta \mathbf{m}_{1}, \delta \mathbf{m}_{2} \right) = \delta \mathbf{m}_{1} \begin{pmatrix} 0 & 2v_{\mathrm{P}} \delta_{ij} \delta_{kl} & 2v_{\mathrm{S}} \gamma_{ijkl} \\ 2v_{\mathrm{P}} \delta_{ij} \delta_{kl} & 2\rho \delta_{ij} \delta_{kl} & 0 \\ 2v_{\mathrm{S}} \gamma_{ijkl} & 0 & 2\rho \gamma_{ijkl} \end{pmatrix} \delta \mathbf{m}_{2},$$
(82)

with the model vector

$$\delta \mathbf{m}_i = (\delta \rho_i, \delta v_{\mathrm{P},i}, \delta v_{\mathrm{S},i}), \quad i = 1, 2,$$
(83)

and the auxiliary variable γ_{ijkl} defined by

$$\gamma_{ijkl} = \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - 2\delta_{ij}\delta_{kl} \,. \tag{84}$$

This allows us to write eq. (80) in a very convenient form

$$\int_{G} K_{m}^{1 \leftrightarrow 1} \,\delta\mathbf{m}_{2} \,d^{3}\mathbf{x} = \int_{G} \delta\mathbf{m}_{1} \mathbf{J} \delta\mathbf{m}_{2} \,d^{3}\mathbf{x} \,, \tag{85}$$

where we defined the symmetric 3×3 matrix **J** through

$$\mathbf{J} = \int_{T} \begin{pmatrix} 0 & 2v_{\mathrm{P}} \operatorname{tr} \boldsymbol{\epsilon}^{\dagger} \operatorname{tr} \boldsymbol{\epsilon} & 4v_{\mathrm{S}} \left(\boldsymbol{\epsilon}^{\dagger} : \boldsymbol{\epsilon} - \operatorname{tr} \boldsymbol{\epsilon}^{\dagger} \operatorname{tr} \boldsymbol{\epsilon} \right) \\ 2v_{\mathrm{P}} \operatorname{tr} \boldsymbol{\epsilon}^{\dagger} \operatorname{tr} \boldsymbol{\epsilon} & 2\rho \operatorname{tr} \boldsymbol{\epsilon}^{\dagger} \operatorname{tr} \boldsymbol{\epsilon} & 0 \\ 4v_{\mathrm{S}} \left(\boldsymbol{\epsilon}^{\dagger} : \boldsymbol{\epsilon} - \operatorname{tr} \boldsymbol{\epsilon}^{\dagger} \operatorname{tr} \boldsymbol{\epsilon} \right) & 0 & 4\rho \left(\boldsymbol{\epsilon}^{\dagger} : \boldsymbol{\epsilon} - \operatorname{tr} \boldsymbol{\epsilon}^{\dagger} \operatorname{tr} \boldsymbol{\epsilon} \right) \end{pmatrix} dt .$$
(86)

From eq. (85) we immediately infer that the Hessian kernel $K_m^{1\leftrightarrow 1}$ is explicitly given by

$$K_m^{1\leftrightarrow 1} = \mathbf{J}\delta\mathbf{m}_1\,.\tag{87}$$

It is interesting to note that the *P* and *S* wave speed Fréchet kernels, K_{v_p} and K_{v_s} , from paragraph (3.2) reappear in the matrix **J**. In fact, we may write **J** in the following form

$$\mathbf{J} = \begin{pmatrix} 0 & \rho^{-1} K_{v_{\rm P}} & \rho^{-1} K_{v_{\rm S}} \\ \rho^{-1} K_{v_{\rm P}} & v_{\rm P}^{-1} K_{v_{\rm P}} & 0 \\ \rho^{-1} K_{v_{\rm S}} & 0 & v_{\rm S}^{-1} K_{v_{\rm S}} \end{pmatrix}$$
(88)

The Fréchet kernels for the isotropic medium can thus be recycled for the computation of the Hessian kernel K_{1}^{Hol} .

3.3.2 Physical interpretation of the Hessian kernels

As mentioned in Section 3.2, Fréchet kernels can be interpreted in terms of first-order scattering from within a primary influence zone (see Fig. 1). Hessian kernels, in contrast, also represent second-order scattering from within secondary influence zones. More specifically, the Hessian kernel $K_m^{1\leftrightarrow 2} = \int_T \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\delta_1 \mathbf{u}) dt$ results from the interaction of the adjoint field \mathbf{u}^{\dagger} with the first-order scattered field

$$\delta_1 \mathbf{u} = \nabla_m \mathbf{u} \,\delta \mathbf{m}_1 = \lim_{\nu \to 0} \frac{1}{\nu} [\mathbf{u}(\mathbf{m} + \nu \,\delta \mathbf{m}_1) - \mathbf{u}(\mathbf{m})], \tag{89}$$

that is excited when the forward field **u** impinges upon the model perturbation $\delta \mathbf{m}_1$. The construction of $K_m^{1\to 2}$ is very similar to the construction of the Fréchet kernel $K_m = \int_T \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\mathbf{u}) dt$ which involves \mathbf{u}^{\dagger} and \mathbf{u} instead of \mathbf{u}^{\dagger} and $\delta_1 \mathbf{u}$. This similarity suggests that the Hessian kernel $K_m^{1\to 2}$ localizes a secondary influence zone where a model perturbation $\delta \mathbf{m}_2$ can cause the first-order scattered field $\delta_1 \mathbf{u}$ to generate a second-order scattered field that affects the observation at the receiver. Fig. 2(a) illustrates this process.



Figure 2. Schematic illustration of the Hessian kernels $K_{\rm m}^{1\rightarrow2}$ and $K_{\rm m}^{2\rightarrow1}$ that are related to first- and second-order scattering. (a) Construction of the Hessian kernel $K_{\rm m}^{1\rightarrow2}$ through the interaction of the the adjoint field \mathbf{u}^{\dagger} and $K_{\rm m}^{2\rightarrow1}$ that are related to first- and second-order scattering. (a) Construction of the Hessian kernel $K_{\rm m}^{1\rightarrow2}$ through the interaction of the the adjoint field \mathbf{u}^{\dagger} and the first-order scattered forward field $\delta_1 \mathbf{u}$ that is generated when the forward field \mathbf{u} impinges on the model perturbation $\delta \mathbf{m}_1$. The Hessian kernel $K_{\rm m}^{1\rightarrow2}$, shown as the shaded area, localizes the volume where a perturbation $\delta \mathbf{m}_2$ causes $\delta_1 \mathbf{u}$ to generate a second-order scattered field that affects the measurement at the receiver (•). (b) The Hessian kernel $K_{\rm m}^{2\rightarrow1}$ is the superposition of two influence zones that correspond to the two sources of the differentiated adjoint field $\delta_1 \mathbf{u}^{\dagger}$. The first source, $-\nabla_{\rm m} \mathbf{L}^{\dagger}(\mathbf{u}^{\dagger}) \delta_{\rm m}_1$, excites a scattered field which then generates a secondary influence zone (left panel). A first-order scattered field that originates from within this secondary influence zone can then be scattered again from $\delta \mathbf{m}_1$ and finally affect the measurement at the receiver. The second source of $\delta_1 \mathbf{u}^{\dagger}$ is $-\nabla_u \nabla_u \chi^{\dagger}(\delta_1 \mathbf{u})$. The corresponding influence zone extends from source to receiver and accounts for the first-order scattering contribution represented by the approximate Hessian \widetilde{H} (right panel).

The interpretation of the Hessian kernel $K_m^{2\to 1} = \int_T \delta_1 \mathbf{u}^{\dagger} \nabla_m \mathbf{L}(\mathbf{u}) dt$ is slightly more involved because the field $\delta_1 \mathbf{u}^{\dagger}$ has two sources, as we have seen already in eq. (46).

One of the two sources, $-\nabla_{\mathbf{m}} \mathbf{L}^{\dagger}(\mathbf{u}^{\dagger}) \delta \mathbf{m}_{1}$, excites a scattered adjoint field that is similar to the scattered forward field $\delta_{1}\mathbf{u}$ but propagates in the opposite direction. The interaction of the scattered adjoint field with the forward field \mathbf{u} generates another secondary influence zone that connects the perturbation $\delta \mathbf{m}_{1}$ to the source. Another perturbation $\delta \mathbf{m}_{2}$ located within this secondary influence zone causes the forward field \mathbf{u} to excite a first-order scattered field $\delta_{2}\mathbf{u}$ which then interacts with the perturbation $\delta \mathbf{m}_{1}$ such that the resulting second-order scattered field affects the measurement made at the receiver. This is shown schematically in the left part of Fig. 2(b).

The second source of $\delta_1 \mathbf{u}^{\dagger}$ is $-\nabla_u \nabla_u \chi^{\dagger}(\delta_1 \mathbf{u})$, which is the derivative of the regular adjoint source. It accounts for the changes of the adjoint source that are due to the perturbation of the forward wavefield \mathbf{u} , and it generates an influence zone that extends from the source to the receiver (Fig. 2b, right-hand side). A comparison with eq. (34) reveals that $-\nabla_u \nabla_u \chi^{\dagger}(\delta_1 \mathbf{u})$ can be interpreted as the source of the approximate Hessian \widetilde{H} which accounts for the first-order scattering contribution to the full Hessian *H*.

The complete Hessian kernel $K_m^{2 \rightarrow 1}$ is thus a superposition of (1) a secondary influence zone that corresponds to second-order scattering, and (2) a primary influence zone that corresponds to the first-order scattering represented by the approximate Hessian.

In the interest of a simplified vocabulary, we will henceforth refer to the wavefield

$$\delta_1 \mathbf{u}^{\dagger} = \nabla_m \mathbf{u}^{\dagger} \,\delta \mathbf{m}_1 = \lim_{\nu \to 0} \frac{1}{\nu} \left[\mathbf{u}^{\dagger} (\mathbf{m} + \nu \,\delta \mathbf{m}_1) - \mathbf{u}^{\dagger} (\mathbf{m}) \right], \tag{90}$$

as the scattered adjoint field, though keeping in mind that it is caused by more than scattering sensu stricto.

The Hessian kernel $K_m^{1\leftrightarrow 1} = \int_T \mathbf{u}^{\dagger} \nabla_m \nabla_m \mathbf{L}(\mathbf{u})(\delta \mathbf{m}_1) dt$ does not appear to have a straightforward intuitive interpretation. It accounts for the non-linearity introduced by the parameterization. Whether $K_{\delta m_1}^{1\leftrightarrow 1}$ is zero or not, depends strongly on the choice of free parameters, as we have seen already in Section 3.3.1.

4 HESSIAN KERNEL GALLERY

To bring the previous developments to life, we continue with specific examples of Hessian kernels for a small selection of data functionals and model perturbations. Given the infinite number of seismic data functionals and Earth model parameterizations, our Hessian kernel gallery can naturally not be exhaustive. We nevertheless hope that it provides both physical intuition and a useful illustration of the methodology outlined in Sections 2 and 3.



Figure 3. Left-hand side: N–S component synthetic seismogram computed using a spectral-element method (Fichtner & Igel 2008). The time window around the Love wave used to compute Fréchet and Hessian kernels is shaded in grey. The dominant period is 25 s. Right-hand side: Fréchet kernel \bar{K}_{vs} for a cross-correlation time-shift measurement on the N–S component Love wave shown to the left-hand side. The source and receiver positions are marked by + and •, respectively.

4.1 Surface wave interaction with point-localized and extended scatterers

We begin with a detailed example where we consider a N–S component Love wave excited by a shallow event at 10 km depth and observed at an epicentral distance of 25.2° (Fig. 3, left-hand side). To quantify the discrepancy between the observed waveform $u^0(t)$ and the synthetic waveform u(t), we measure the cross-correlation time shift T, defined as the global maximum of the correlation function

$$\mathcal{C}(u^0, u)(t) = \int_T u^0(\tau) u(\tau + t) d\tau \,. \tag{91}$$

We have T > 0 when the synthetic waveform arrives later than the observed waveform, and T < 0 when the synthetic waveform arrives earlier than the observed waveform. Under the assumption that $u^0(t)$ and u(t) are merely shifted in time without being otherwise distorted with respect to each other, the adjoint source corresponding to the data functional $\mathcal{X} = T$ is given by (Luo & Schuster 1991)

$$\mathbf{f}^{\dagger}(\mathbf{x},t) = \frac{\mathbf{e}_{\mathrm{NS}}}{||\dot{u}||_{2}^{2}} \dot{u}(t) \,\delta(\mathbf{x} - \mathbf{x}^{r}) \,, \tag{92}$$

where \mathbf{e}_{NS} and \mathbf{x}^r denote the unit vector in N–S direction and the receiver location, respectively. The normalization by the squared L_2 norm $||\dot{u}||_2^2 = \int_T \dot{u}^2(t) dt$ ensures that the resulting Fréchet kernels do not depend on amplitude. Treating observed and synthetic waveforms as time-shifted versions of each other is a simplified but widely used approach because it leads to kernels that are quasi-independent of the actual data. We can thus compute kernels without explicitly measuring a time shift.

The computation of the Fréchet kernels follows a well-known recipe. First, we solve the forward problem (eqs 62 and 63) to obtain the forward field **u** that is stored at sufficiently many time steps. Snapshots of **u**, computed with the help of a spectral-element discretization of the seismic wave equation (Fichtner & Igel 2008), are shown in the left panel of Fig. 4. We then select a time window around the waveform of interest (Fig. 3, left-hand side) and compute the adjoint source given in eq. (92). The adjoint source \mathbf{f}^{\dagger} is the right-hand side of the adjoint eq. (21), the solution of which is the adjoint field \mathbf{u}^{\dagger} , shown in the left panel of Fig. 5. While the adjoint field propagates in reverse time, the previously stored forward field is loaded and used to compute the Fréchet kernels according to eqs (76) and (77).

The S velocity kernel \bar{K}_{v_S} corresponding to the cross-correlation time shift measurement on the N–S component Love wave is shown in the right panel of Fig. 3. The kernel is dominated by negative sensitivities within the first Fresnel zone, where a positive v_S perturbation leads to earlier-arriving waveforms and to a decrease in the cross-correlation time shift T. The width of the first Fresnel zone is proportional to $\sqrt{T_d \ell}$, where T_d and ℓ denote the dominant period and the length of the ray path between source and receiver, respectively.

The Hessian kernel $K_m^{1\leftrightarrow 1}$ can be computed directly from the Fréchet kernels (eq. 88). Additional simulations are needed for $K_m^{1\rightarrow 2}$ and $K_m^{2\rightarrow 1}$ because these involve the scattered forward field $\delta_1 \mathbf{u}$ and the scattered adjoint field $\delta_1 \mathbf{u}^{\dagger}$ (eqs 58 and 59). Both scattered fields can be computed most conveniently with the help of the finite-difference approximations

$$\delta_1 \mathbf{u} = \nabla_m \mathbf{u} \,\delta \mathbf{m}_1 \approx \frac{1}{\nu} [\mathbf{u}(\mathbf{m} + \nu \,\delta \mathbf{m}_1) - \mathbf{u}(\mathbf{m})], \tag{93}$$

and

$$\delta_1 \mathbf{u}^{\dagger} = \nabla_m \mathbf{u}^{\dagger} \,\delta \mathbf{m}_1 \approx \frac{1}{\nu} \left[\mathbf{u}^{\dagger} (\mathbf{m} + \nu \,\delta \mathbf{m}_1) - \mathbf{u}^{\dagger} (\mathbf{m}) \right], \tag{94}$$

where v > 0 is a sufficiently small scalar. In our example, we choose $\delta \mathbf{m}_1$ to be a 10 per cent v_s perturbation, δv_{s1} , located within a 50 × 50 km³ box-shaped basis function right beneath the surface at nearly equal distance from the source and the receiver (Figs 4 and 5). The v_s perturbation almost acts as a point scatterer because its dimension is small compared to the dominant wavelength of around 130 km. We are therefore close to the Rayleigh scattering regime. The scattered forward field $\delta_1 \mathbf{u}$ is excited when the regular wave front reaches δv_{s1} which acts as a fictitious source. It then propagates towards the receiver (Fig. 4, right-hand side). In contrast, the pure scattering contribution to $\delta_1 \mathbf{u}^{\dagger}$ collapses into the v_s perturbation δv_{s1} (Fig. 5).

The resulting Hessian kernels $\bar{K}_{v_{s}}^{2 \to 1}$, $\bar{K}_{v_{s}}^{1 \to 1}$ and $\bar{K}_{v_{s}}^{1 \to 2}$ are shown separately in the top row of Fig. 6. The kernel $\bar{K}_{v_{s}}^{2 \to 1}$ is a superposition of two contributions, labelled *F* and *S*. These correspond to the primary influence zone represented by the approximate Hessian (*F*) and to



Figure 4. N–S component of the forward wavefield **u** (left column) and the scattered forward wavefield δ_1 **u** plotted at the surface. The source is marked by + and the receiver by • (epicentral distance =25.2°). The location of the v_S perturbation is indicated by **I**. The images are dominated by surface waves. Small-amplitude body waves are hardly visible. The v_S perturbation acts as a point scatterer (Rayleigh scattering), and most of the Love wave energy is scattered in the forward direction.

the secondary influence zone where second-order scattering affects the measurement (See Fig. 2 for a schematic representation). The kernel $\bar{K}_{v_{\rm S}}^{1\to2}$ only extends from the receiver to δv_{S1} . Second-order scattering from δv_{S1} to an *S* velocity perturbation within $\bar{K}_{v_{\rm S}}^{1\to2}$ has an effect on the measurement. The kernel $\bar{K}_{v_{\rm S}}^{1\leftrightarrow1}$ is restricted to the volume occupied by δv_{S1} . The complete Hessian kernel $\bar{K}_{v_{\rm S}}^{1} = \bar{K}_{v_{\rm S}}^{2\to1} + \bar{K}_{v_{\rm S}}^{1\to2} + \bar{K}_{v_{\rm S}}^{1\to2}$ is shown in the bottom row of Fig. 6.

An interesting observation is that the sign of $\bar{K}_{v_{S}}^{1}$ within the secondary influence zones is opposite (positive) to the sign within the primary influence zone (negative) that corresponds to the approximate Hessian. The widths of the secondary influence zones is proportional to $\sqrt{T_{d}\ell_{s\rightarrow 1}}$ and $\sqrt{T_{d}\ell_{1\rightarrow r}}$, where the $\ell_{s\rightarrow 1}$ and $\ell_{1\rightarrow r}$ signify the length of the ray paths from the source to δv_{S1} and from δv_{S1} to the receiver, respectively. This explains why the secondary influence zones are slim compared to the primary influence zone (labelled *F* in Fig. 6) where the width is proportional to $\sqrt{T_{d}(\ell_{s\rightarrow 1} + \ell_{1\rightarrow r})}$.

The Hesian kernel K_m^1 shown in Fig. 6 can be interpreted as the continuous representation of the row of the Hessian matrix **H** that corresponds to the basis function coincident with δv_{S1} . The fully discrete row is obtained by projecting K_m^1 onto the basis functions, according to eq. (61).

The amplitude of the Hessian kernel $\bar{K}_{v_{S}}^{1}$ for δv_{S1} located near the centre of the first Fresnel zone is about two orders of magnitude smaller than the amplitude of the Fréchet kernel. This suggests that the single-scattering approximation is well justified in this particular scenario. Locating δv_{S1} further away from the first Fresnel zone rapidly decreases the amplitude of the corresponding Hessian kernels, as can be seen in Fig. 7. The small second derivatives result from the long propagation distance of the scattered waves that arrive too late to have a significant effect inside the Love wave time window (Fig. 3, left-hand side).

The recipe for the computation of Hessian kernels is equally applicable when the model perturbation (or basis function) does not effectively act as a point scatterer, as in the previous example. For a spatially extended model perturbation δv_{S1} , such as the one shown in Fig. 8, we are in the Mie scattering regime where the characteristic size of the scatterer is much larger than the dominant period. To calculate the Hessian kernel corresponding to δv_{S1} , we again compute the scattered fields $\delta_1 \mathbf{u}$ and $\delta_1 \mathbf{u}^{\dagger}$ using the finite-difference approximations from eqs (93) and (94), respectively. The finite-difference step length ν must be sufficiently small to ensure the convergence of the approximation. For the specific perturbation in Fig. 8, we found $\nu = 1/100$ to be appropriate.

Slices through the resulting Hessian kernel K_{vs}^1 are shown in Fig. 9. As expected for Love waves, the Hessian kernel is largest near the surface and decays rapidly with increasing depth. An interesting observation is the strong asymmetry of the Hessian kernel, as compared to the one computed for an effective point scatterer (Fig. 6). The origin of the asymmetry is difficult to identify on the basis of purely numerical simulations, though we hypothesize that it is related to the non-isotropic radiation pattern of the regular source.



Figure 5. N–S component of the adjoint wavefield \mathbf{u}^{\dagger} (left column) and the pure scattering contribution to $\delta_1 \mathbf{u}^{\dagger}$ (disregarding the contribution to $\delta_1 \mathbf{u}^{\dagger}$ from the change of the adjoint source) plotted at the surface. The adjoint wavefield emanates from the receiver (•) and propagates in reverse time towards the source (+). The adjoint source corresponds to the measurement of a cross-correlation time-shift of the N–S component Love wave at a dominant period of 25 s. The location of the v_S perturbation is indicated by \blacksquare . Most of the Love wave energy is scattered forward, that is, in the propagation direction of the adjoint field. The contribution to $\delta_1 \mathbf{u}^{\dagger}$ that comes from the perturbation of the adjoint source (see eq. 34 and Section 3.3.2) is not shown to enhance the visibility of the pure scattered wave.

4.2 P body waves and the second-order effect of density

The application of the previously described methodology to other types of seismic waves is straightforward and merely requires the choice of different time windows. In the following example, we consider the measurement of a cross-correlation time shift on the vertical component of an 8 s *P* body wave in the same source–receiver geometry as above. We furthermore extend the analysis to the multiparameter case. The Fréchet kernels for v_P , v_S and ρ are displayed in Fig. 10(a). As expected, the *P* wave traveltime is mostly affected by perturbations in v_P , whereas variations in v_S and ρ have hardly any first-order effect.

To study second-order effects on the finite-frequency traveltime of the *P* wave, we again place a $50 \times 50 \times 50 \text{ km}^3 P$ velocity perturbation, δv_{P1} , in the centre of the first-order first Fresnel zone of the v_P Fréchet kernel \bar{K}_{v_P} . The dimension of δv_{P1} is small compared to the wavelength (~100 km), meaning that we are close to the Rayleigh scattering regime. The corresponding Hessian kernels, shown in Fig. 10b, contain two symmetric lobes that meet at the location of the scatterer. These lobes correspond to the second-order scattering contribution to the Hessian, and are labelled *S* in the $K_{v_P}^1$ kernel. The contribution from the approximate Hessian, labelled *F*, is clearly visible only in $K_{v_P}^1$. The Hessian kernels $K_{v_S}^1$ and K_{ρ}^1 are small compared to $K_{v_P}^1$. This indicates that second-order scattering from perturbations of different nature, that is, one perturbation in v_P and another one in ρ or v_S , is rather inefficient.

For a scatterer located outside the first Fresnel zone, we observe a similar decrease of the Hessian kernels as for surface waves (see Fig. 7). In the specific case of a v_P perturbation near the surface, the corresponding Hessian kernels are around two orders of magnitude smaller than the kernels for δv_{P1} located in the centre of the Fréchet kernel (Fig. 10c). The fact that scattering from near the surface affects the traveltime measurement at all, can be explained with the finite-frequency content of the *P* waveform. Doubly-scattered energy from δv_{P1} and another v_P perturbation located within the support of $\bar{K}_{v_P}^1$ can still arrive in the measurement window of the *P* wave that is around 15 s long.

The kernels shown in Figs 10(a) and (b) reveal that density perturbations have hardly any effect of the finite-frequency traveltime of P waves through first-order scattering or second-order scattering that also involves a P velocity perturbation. To see whether second-order scattering from two density perturbations can influence the P wave traveltime, we repeat the example from Fig. 10(b), but we replace δv_{P1} by a density scatterer $\delta \rho_1$. The resulting Hessian kernels, shown in Fig. 10(d), are practically zero, thus indicating that variations in density do not have any significant second-order effect on P wave traveltimes at all.

This observation is closely related to the characteristics of Rayleigh scattering from density perturbations (e.g. Wu & Aki 1985; Tarantola 1986). For $\delta \rho_1 \neq 0$ and $\delta v_{P1} = \delta v_{S1} = 0$ most of the energy is scattered in the direction opposite to the propagation direction of the incident wave (Fig. 11). Therefore, the scattered forward field $\delta_1 \mathbf{u}$ does not interact with the adjoint field \mathbf{u}^{\dagger} , and the partial Hessian kernel $K_m^{1\rightarrow 2}$ is close to zero. Similarly, the scattered adjoint field $\delta_1 \mathbf{u}^{\dagger}$ propagates in a direction opposite to the adjoint field \mathbf{u}^{\dagger} , so that the interaction of



Figure 6. Hessian kernels plotted at the surface of the Earth. The position δv_{S1} is indicated by \circ . Top row: The individual Hessian kernels $\bar{K}_{vS}^{2\to1}$, $\bar{K}_{vS}^{1\to1}$ and $\bar{K}_{vS}^{1\to2}$ (from left-hand side to right-hand side). The kernel $\bar{K}_{vS}^{2\to1}$ has two contributions, labelled *F* and *S*. Contribution *F* represents the primary influence zone that corresponds to the approximate Hessian. Contribution *S*, encircled by a dotted curve, is a secondary influence zone where second-order scattering affects the measurement. (In Section 5.1, we show how the approximate Hessian can be computed separately.) Bottom row: Composite Hessian kernel \bar{K}_{vS}^1 . The the contribution from second-order scattering is again encircled by a dotted curve and marked with *S*.



Figure 7. Hessian kernel $\bar{K}_{v_{\rm S}}^1$ corresponding to an *S* velocity perturbation δv_{S1} that is located in the outer rim of the Fréchet kernel shown in Fig. 3. The position δv_{S1} is indicated by \circ . The amplitude of $\bar{K}_{v_{\rm S}}^1$ is one order of magnitude smaller than in Fig. 6 where δv_{S1} was located near the centre of the first Fresnel zone of the Fréchet kernel.

 $\delta_1 \mathbf{u}^{\dagger}$ with the forward field \mathbf{u} is not possible. It follows that $K_m^{2 \to 1}$ is approximately zero as well. Finally, we have $K_m^{1 \leftrightarrow 1} = 0$ for pure density perturbations (eq. 88), and the complete Hessian kernel $K_m^1 = K_m^{2 \to 1} + K_m^{1 \leftrightarrow 1} + K_m^{1 \to 2}$ nearly vanishes.

5 FULL HESSIAN VERSUS APPROXIMATE HESSIAN

The approximate Hessian \tilde{H} is commonly used as a computationally less expensive substitute of the full Hessian H, because its computation merely requires first derivatives. To explore the potential differences between \tilde{H} and H, we consider—as an example—a realistic 3-D full waveform tomography. As a preparatory step, however, we expand on the computation of approximate Hessian kernels, using the methodology developed in Sections 2 and 3.

5.1 Computing approximate Hessian kernels

The approximate Hessian is defined as that part of the full Hessian that does not contain second derivatives (see eq. 34)

$$H(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = \langle \nabla_u \nabla_u \chi(\delta_1 \mathbf{u}, \delta_2 \mathbf{u}) \rangle = \langle \delta_2 \mathbf{u} \nabla_u \nabla_u \chi^{\mathsf{T}}(\delta_1 \mathbf{u}) \rangle .$$

Our goal is again to isolate the second model perturbation $\delta \mathbf{m}_2$. For this, we define an approximate scattered adjoint field $\delta_1 \mathbf{\tilde{u}}^{\dagger}$ as the solution of the adjoint equation

$$\mathbf{L}^{\dagger}(\delta_{1}\widetilde{\mathbf{u}}^{\dagger}) = -\nabla_{u}\nabla_{u}\chi^{\dagger}(\delta_{1}\mathbf{u}).$$
⁽⁹⁶⁾

© 2011 The Authors, *GJI*, **185**, 775–798 Geophysical Journal International © 2011 RAS

(95)



Figure 8. Horizontal (top) and vertical (bottom) slice through the extended v_S perturbation used to compute the Hessian kernel $\bar{K}_{v_S}^1$ shown in Fig. 9. The maximum perturbation of 0.4 km s⁻¹ near the surface is equal to 10 per cent of the background *S* velocity.



Figure 9. Horizontal (left-hand side) and vertical (right-hand side) slices through the Hessian kernel \bar{K}_{vs}^1 that corresponds to the v_s perturbation shown in Fig. 8.

The comparison of (96) with eq. (46) reveals that $\delta_1 \tilde{\mathbf{u}}^{\dagger}$ is the contribution to the scattered adjoint field $\delta_1 \mathbf{u}^{\dagger}$ that is excited only by the change of the adjoint source. Inserting (96) into (95) yields

$$\widetilde{H}(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = -\left\langle \delta_2 \mathbf{u} \mathbf{L}^{\dagger}(\delta_1 \widetilde{\mathbf{u}}^{\dagger}) \right\rangle = -\left\langle \delta_1 \widetilde{\mathbf{u}}^{\dagger} \mathbf{L}(\delta_2 \mathbf{u}) \right\rangle.$$
(97)

Then, upon using eq. (48), we can rewrite (97) as

$$\widetilde{H}(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = \left\langle \delta_1 \widetilde{\mathbf{u}}^{\dagger} \nabla_m \mathbf{L}(\mathbf{u}) \, \delta \mathbf{m}_2 \right\rangle \,. \tag{98}$$

From (98) we finally infer that the approximate Hessian kernel \widetilde{K}_m^1 is given by

$$\widetilde{K}_{m}^{1} = \int_{T} \delta_{1} \widetilde{\mathbf{u}}^{\dagger} \nabla_{m} \mathbf{L}(\mathbf{u}) dt .$$
⁽⁹⁹⁾

With eq. (99) in mind we now delve into our example that allows us to study the potential differences between the approximate and full Hessians in a realistic application.

5.2 A realistic example (Influences on our perception of the Iceland plume)

We consider a long-period full waveform tomography for the European upper mantle that is summarized in Fig. 12. The data used in the inversion are three-component seismograms with a dominant period of 100 s, that provide a good coverage of central and northern Europe (Fig. 12, left-hand side). The inversion was based on the measurement of time- and frequency-dependent phase differences (Fichtner *et al.* 2008). As initial model we used the 3-D mantle structure from S20RTS (Ritsema *et al.* 1999) combined with the crustal model by Meier *et al.* (2007a,b). After three conjugate-gradient iterations we obtained the tomographic images that are shown in the centre and right panels of Fig. 12, in the form of perturbations relative to the 1-D model AK135 (Kennett *et al.* 1995). The achieved waveform fit and the reduction of the initial misfit by nearly 70 per cent indicate that the tomographic model is at least close to optimal.

One of the most prominent features in the images is the Iceland plume (Fig. 12, right-hand side) that we choose as our model perturbation δv_{S1} . Horizontal slices through the isolated Iceland plume are shown in the left column of Fig. 13. Then, following the methodology from Sections 2, 3 and 5.1, we compute the corresponding approximate and full Hessian kernels, displayed in the centre and right columns of Fig. 13, respectively.

792 A. Fichtner and J. Trampert

a) Fréchet kernels



b) Hessian kernels: δvp>0, δvs=0, δρ=0 [scatterer within the 1st Fresnel zone]



c) Hessian kernels: δvp>0, δvs=0, δp=0 [surface scatterer]



d) Hessian kernels: δvp=0, δvs=0, δρ>0 [scatterer within the 1st Fresnel zone]



Figure 10. Fréchet and Hessian kernel gallery for the cross-correlation time-shift on the vertical component of an 8 s P wave. The source-receiver geometry is the same as in Fig. 3. (a) Fréchet kernels for v_P , v_S and ρ (from left-hand side to right-hand side). (b) Hessian kernels $\bar{K}^1_{v_P}$, $\bar{K}^1_{v_S}$ and \bar{K}^1_{ρ} (from left-hand side to right-hand side). (c) Hessian kernels $\bar{K}^1_{v_P}$, $\bar{K}^1_{v_S}$ and \bar{K}^1_{ρ} (from left-hand side to right-hand side). (c) Hessian kernels $\bar{K}^1_{v_P}$, $\bar{K}^1_{v_S}$ and \bar{K}^1_{ρ} (from left-hand side to right-hand side) for a v_P perturbation δv_{P1} . The position of δv_{P1} is indicated by \circ . In the $\bar{K}^1_{v_P}$ kernel the contributions from the second-order scattering and the approximate Hessian are outlined by dash-dotted curves and labelled S and F, respectively. The amplitudes of $\bar{K}^1_{v_S}$ and \bar{K}^1_{ρ} are small compared to the amplitude of $\bar{K}^1_{v_P}$. (c) The same as in (b) but for a v_P perturbation located directly beneath the surface (\circ). The kernels are comparatively small but different from zero, indicating that second-order scattering from far outside the Fréchet kernel may affect the measurement because the P wave window is sufficiently long (\sim 15 s). (d) Hessian kernels $\bar{K}^1_{v_P}$, $\bar{K}^1_{v_S}$ and \bar{K}^1_{ρ} (from left-hand side to right-hand side) for a density perturbation $\delta \rho_1$. The position of $\delta \rho_1$ is indicated by \circ . The kernels are characterized by rapid variations that are localized around the density perturbation. Compared to the Hessian kernels for a v_P perturbation in (b), the density Hessian kernels have small amplitudes, indicating that ρ has hardly any second-order effect on the finite-frequency traveltime of P waves.

Each of the kernels can be interpreted as the continuous representation of that row in the (approximate or full) Hessian that is associated with the Iceland plume. The volume occupied by the plume itself, and indicated by dashed circles in Fig. 13, represents the diagonal element. Off-diagonal elements correspond to volumes outside the plume.

Using this terminology, we can say that the diagonal elements of the approximate and full Hessian kernels are nearly identical at depths of around 100 km. This similarity, however, decreases steadily with increasing depth. Large differences in the off-diagonal elements, including opposite signs, are most pronounced beneath Greenland and Central Europe. This indicates that the second-order scattering contribution to the Hessian may not be negligible, even in the vicinity of the optimal model.

The Hessian kernels from Fig. 13 already provide semi-quantitative insight into where and how our image of the Iceland plume is affected by v_S structure elsewhere in the upper mantle. While structure beneath southern and eastern Europe has no effect on our perception of the plume ($K_{v_S}^1 = 0$), its average v_S is dependent on v_S beneath northern Europe and parts of Greenland ($K_{v_S}^1 \neq 0$). Based on the quadratic approximation of the misfit functional (eq. 3) we infer that increasing v_S in regions where $K_{v_S}^1 < 0$ can partly be compensated by decreasing v_S within the plume volume, and vice versa. Note that the approximate Hessian kernel $\tilde{K}_{v_S}^1$ would lead us to the erroneous conclusion that a decrease (increase) of v_S beneath Greenland combined with a decrease (increase) of the average plume v_S would have little effect on the misfit \mathcal{X} .



Figure 11. Vertical slices through the E–W (left-hand side), N–S (centre) and Z components of the scattered velocity field $\delta \dot{\mathbf{u}}$. The snapshots are taken 30 s after the *P* wave reached the density perturbation $\delta \rho_1$, marked by \circ . The direction of the incident *P* wave is indicated by left-pointing arrows. Clearly visible is the backward scattering associated with $\delta \rho_1 > 0$ and $\delta v_{S1} = \delta v_{P1} = 0$. Hardly any energy is scattered in the propagation direction of the regular wavefield \mathbf{u} .



Figure 12. Full waveform tomography for the European upper mantle. Left-hand side: Ray coverage. Centre: Relative *S* wave speed perturbations at 100 km depth. Right-hand side: Zoom on the Iceland hotspot at various depth levels. The reference model is AK135 (Kennett *et al.* 1995).

6 DISCUSSION

In the previous sections, we presented an extension of the well-known adjoint method that allows us to compute the second derivatives of seismic data functionals with respect to Earth model parameters. The Hessian applied to a model perturbation, $H\delta \mathbf{m}$, can be represented by Hessian kernels that are similar to Fréchet kernels but also involve scattered forward and adjoint fields. This work is intended to serve as a technical prelude for the implementation of Newton's method and the development of quantitative resolution analyses in full waveform inversion. In the following paragraphs, we discuss issues related to computational aspects and to the interpretation of the Hessian kernels.

6.1 Computational requirements and the practical usefulness of Hessian kernels

The computing time required for the calculation of Hessian kernels is twice as long as for Fréchet kernels, because four instead of two wavefield simulations are needed. First, the regular wavefield **u** has to be modelled and stored at sufficiently many intermediate time steps. The Fréchet kernel K_m and the partial Hessian kernel $K_m^{1\leftrightarrow 1}$ can be computed during the adjoint simulation from the interaction of **u** and the adjoint field \mathbf{u}^{\dagger} . Two additional simulations are then needed for the perturbed regular and adjoint fields, $\mathbf{u}(\mathbf{m} + \nu \delta \mathbf{m}_1)$ and $\mathbf{u}^{\dagger}(\mathbf{m} + \nu \delta \mathbf{m}_1)$, that enter the finite-difference approximations for $\delta_1 \mathbf{u}$ and $\delta_1 \mathbf{u}^{\dagger}$, respectively (eqs 93 and 94).

The comparatively high numerical costs involved in the computation of $H\delta \mathbf{m}_1 \equiv K_m^1$ naturally constrain the practical usefulness of the Hessian kernels. The competitiveness of Newton's method in particular will depend critically on the development of efficient algorithms that yield good approximate solutions of Newton's eq. (2) with as few evaluations of $H\delta \mathbf{m}$ as possible. At least for 1-D full waveform inversion, a full Newton method has already proven efficient (Santosa & Symes 1988; Pratt *et al.* 1998).

When second derivatives are used for resolution analysis, then we are in a more favourable situation. This is because the Hessian need not be inverted to obtain a quadratic approximation of the misfit functional \mathcal{X} (eq. 3). The second-order expansion of \mathcal{X} about the optimal model already provides invaluable information about resolution and trade-offs. By choosing model perturbations $\delta \mathbf{m}$ to coincide with well-defined geological objects, the Hessian kernel formalism can be used to quantify their effect on the misfit as well as mutual dependencies (Fichtner 2010). Only for a covariance analysis sensu stricto, the inverse Hessian needs to be approximated iteratively.



Figure 13. Comparison of approximate and full Hessian kernels. Left-hand side column: Slices through the Iceland plume, which serves as model perturbation $\delta \mathbf{m}_1 = (0, 0, \delta v_{S1})$. Central column: Slices through the approximate Hessian kernel $\widetilde{K}_{v_S}^1$. The dashed circle indicates the approximate location of the model perturbation. Right-hand side column: Slices through the full Hessian kernel $K_{v_S}^1$. The colour scales for the approximate and full Hessian kernels are the same at each depth level.

6.2 Inversion for density?

Our seismically inferred knowledge on density structure is comparatively poor and almost exclusively based on the gravitational effect on long-period free oscillations (Kennett 1998; Ishii & Tromp 2001, 2004; Resovsky & Trampert 2003; Trampert *et al.* 2004). This is because the sensitivity of traveltime observations to density is practically zero (see for instance Fig. 10a). However, 2-D and 3-D synthetic full waveform inversions suggest that variations in ρ may also be constrained by short-period data where gravity is negligible (Köhn *et al.* (2010); Y. Capdeville, private communication, 2008). Our hope was therefore that density structure may have a second-order effect on traveltimes. Unfortunately, this does not seem to be the case (Figs 10b and d).

While one may argue that traveltime is not an adequate measurement to detect density variations, the scattering characteristics shown in Fig. 11 suggest that the invisibility of ρ is a more general phenomenon, at least in a transmission tomography setting. Modifications of the measurement will affect the details of the adjoint field. Yet, the scattering direction of the regular and adjoint fields from a pure density perturbation with $\delta v_S = \delta v_P = 0$ will remain to be opposite to the respective incidence directions. It follows that the second derivatives of any transmission tomography data functional with respect to density are small, provided that the Earth model is parameterized in terms of v_P , v_S and ρ . Changing the parameterization for instance to $\mathbf{m} = (\kappa, \mu, \rho)$, will cause the density derivatives to become alive, though at the expense of substantial trade-offs between the parameters.

6.3 Full Hessian versus approximate Hessian

As we have seen in Section 5, the approximate Hessian \hat{H} can differ substantially from the full Hessian H, even when the model is close to optimal. The differences include sign changes and are most pronounced in the off-diagonal elements.

The explanation for the discrepancy between \hat{H} and H is contained in eq. (6), and it has two components. (1) The inherent non-linearity of the phase difference measurements and (2) The remaining waveform misfit that is non-zero despite the near-optimality of the model. Both factors appear to be very general because realistic misfits are never zero, and because any measurement has a non-linear contribution. Therefore, we hypothesise that the approximate and full Hessians will be different in other applications as well. This conjecture is supported by the results of Santosa & Symes (1988), who conducted 1-D synthetic waveform inversions.

The extent to which the difference between \hat{H} and H is practically relevant, depends strongly on the particular application. The approximate Hessian is likely to be more efficient when only the diagonal elements are used to pre-condition a descent direction within an iterative optimization scheme. However, in Newton-like methods that include the off-diagonal elements, the full Hessian may effectively lead to faster convergence. To the best of our knowledge, there is so far no experience concerning this issue. A quantitative comparison between conjugate-gradient (e.g. Mora 1987, 1988; Tape *et al.* 2007; Fichtner *et al.* 2009), quasi-Newton (e.g. Liu & Nocedal 1989; Pratt *et al.* 1998; Epanomeritakis *et al.* 2008; Brossier *et al.* 2009) and full Newton methods for realistic large-scale problems remains to be done.

In the context of resolution analysis, the full Hessian should be used because the approximate Hessian can lead to erroneous inferences concerning trade-offs between model parameters, as we have seen in the example from Fig. 13. Only the full Hessian allows us to correctly account for the effect of non-linearity on model resolution.

6.4 Amplitude of traveltime Hessian kernels

Fréchet kernels describe the first-order effect of structure on the measurement. For instance, a positive v_P anomaly placed within the first Fresnel zone of the Fréchet kernel in Fig. 10(a) decreases the *P* wave traveltime shift \mathcal{T} because the synthetic *P* wave arrives earlier. The influence of the second-order effect is less clear because the Hessian kernels contain both strong negative and positive contributions. In any case, the physical relevance of the second-order effect of course needs to be considered relative to both the first-order effect and the unknown contributions of orders three and higher. A comparison of the Fréchet and Hessian kernels shown in Figs 3, 6 and 9 reveals, that the second-order effect on traveltimes can be nearly as large as the first-order effect, provided that the extent of the heterogeneity is sufficiently large. From our experience, however, we know that traveltime is quasi-linearly related to large-scale Earth structure (e.g. Tong *et al.* 1998). This suggests that the cumulative contribution of all higher-order terms may to some extent compensate for the non-linearity that the second derivatives appear to suggest.

7 CONCLUSIONS

We presented an extension of the well-known adjoint method to the computation of the Hessian applied to a model perturbation. The calculations involve the forward wavefield and the adjoint wavefield, as well as their scattered versions. Our approach naturally leads to the concept of Hessian kernels that can be interpreted as the continuous representation of rows (or columns) in the discrete Hessian matrix. The Hessian kernels appear as the superposition of (1) a first-order influence zone that represents the approximate Hessian and (2) second-order influence zones that represent second-order scattering. The Hessian kernels can be represented in terms of Fréchet kernels, which allows for their easy computation using codes with pre-existing adjoint capabilities.

In a series of examples, we examined second-order effects on finite-frequency traveltimes of both surface and body waves. From this we draw the following conclusions: (1) The second-order effect is relevant only when heterogeneities are located within the first Fresnel zone. This is consistent with ray-theory. (2) Second-order scattering that involves heterogeneities of different nature (e.g. v_P and v_S perturbations) appears to be inefficient, at least for P waves. (3) Second derivatives of any data functional with respect to density are nearly zero, provided that the model is parameterized in terms of density and seismic wave speeds. This result can be explained with the backward-scattering from density perturbations.

Based on a realistic full waveform tomography for European upper-mantle structure, we have shown that significant differences exist between the approximate Hessian and the full Hessian, even in the vicinity of the optimal model. These differences are largest for the off-diagonal elements, and most relevant in resolution and trade-off analysis.

ACKNOWLEDGMENTS

First and most of all we would like to thank Theo van Zessen for maintaining the HPC cluster at the Department of Earth Sciences at Utrecht University. Furthermore, we are grateful to Christian Böhm for many vivid discussions on non-linear optimization and the computation of the Hessian for seismic observables. Insightful comments by Brian Kennett, Guust Nolet, Jeroen Ritsema, Qinya Liu, Moritz Bernauer and two anonymous reviewers helped us to improve the manuscript substantially. The numerical simulations presented in this paper were performed on the HPC system HUYGENS, and funded by NCF/NWO under grant SH-161-09.

REFERENCES

- Aki, K. & Richards, P., 2002. *Quantitative Seismology*, University Science Books, Sausalito, CA.
- Babuška, V. & Cara, M., 1991. Seismic Anisotropy in the Earth, Kluwer Academic Publishers, Dordrecht, Boston, London.
- Baig, A.M. & Dahlen, F.A., 2004. Traveltime biases in random media and the s-wave discrepancy, *Geophys. J. Int.*, 158, 922–938.
- Bamberger, A., Chavent, G. & Lailly, P., 1977. Une application de la théorie du contrôle à un problème inverse sismique, *Ann. Geophys.*, 33, 183– 200.
- Bamberger, A., Chavent, G., Hemons, C. & Lailly, P., 1982. Inversion of normal incidence seismograms, *Geophysics*, 47, 757–770.
- Bleibinhaus, F., Hole, J.A. & Ryberg, T., 2007. Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging, *J. geophys. Res.*, **112**, doi:10.1029/2006JB004611.
- Bleibinhaus, F., Lester, R.W. & Hole, J.A., 2009. Applying waveform inversion to wide-angle seismic surveys, *Tectonophysics*, 472, 238–248.
- Brossier, R., Operto, S. & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, 74, WCC105–WCC118.
- Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances, *Geophys. J. Int.*, submitted.
- Chen, P., Zhao, L. & Jordan, T.H., 2007. Full 3d tomography for the crustal structure of the los angeles region, *Bull. seism. Soc. Am.*, 97, 1094– 1120.
- Dahlen, F.A. & Tromp, J., 1998, *Theoretical Global Seismology*, Princeton University Press, Princeton, NJ.
- Dessa, J.X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Virieux, J. & Kaneda, Y., 2004. Multiscale seismic imaging of the eastern nankai trough by full waveform inversion, *Geophys. Res. Lett.*, **31**, doi:10.1029/2004GL020453.
- Devaney, A.J., 1984. Geophysical diffraction tomography, *IEEE Trans. Geosci. Remote Sens.*, 22, 3–13.
- Epanomeritakis, I., Akcelik, V., Ghattas, O. & Bielak, J., 2008. A newton-cg method for large-scale three-dimensional elastic full waveform seismic inversion, *Inverse Probl.*, 24, doi:10.1088/0266–5611/24/3/034015.
- Fichtner, A., 2010, Full Seismic Waveform Modelling and Inversion. Springer, Heidelberg.
- Fichtner, A. & Igel, H., 2008. Efficient numerical surface wave propagation through the optimization of discrete crustal models—a technique based on non-linear dispersion curve matching (DCM), *Geophys. J. Int.*, **173**, 519–533.
- Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology–I. Theory, *Phys. Earth planet. Inter.*, **157**, 86–104.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.*, 175, 665–685.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**, 1703–1725.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and pasts states of the Australasian upper mantle, *Earth planet. Sci. Lett.*, 290, 270–280.
- Friederich, W., 1999. Propagation of seismic shear and surface waves in a laterally heterogeneous mantle by multiple forward scattering, *Geophys. J. Int.*, **136**, 180–204.
- Gao, F.C., Levander, A.R., Pratt, R.G., Zelt, C.A. & Fradelizio, G.L., 2006. Waveform tomography at a groundwater contamination site, *Geophysics*, 71, H1–H11.
- Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics*, 51, 1387–1403.
- Griewank, A. & Walther, A., 2000. An implementation of checkpointing for the reverse or adjoint mode of computational differentiation, *Trans. Math. Softw.*, **26**, 19–45.

- Hinze, M., Pinnau, R., Ulbrich, M. & Ulbrich, S., 2009, Optimization with PDE Constraints, Springer, Heidelberg.
- Igel, H., Djikpesse, H. & Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, *Geophys. J. Int.*, **124**, 363–371.
- Ishii, M. & Tromp, J., 2001. Even-degree lateral variations in the Earth's mantle constrained by free oscillations and the free-air gravity anomaly, *Geophys. J. Int.*, 145, 77–96.
- Ishii, M. & Tromp, J., 2004. Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive normal modes, *Phys. Earth planet. Inter.*, 146, 113–124.
- Kennett, B.L.N., 1998. On the density distribution within the Earth, Geophys. J. Int., 132, 374–382.
- Kennett, B.L.N., 2001, *The Seismic Wavefield I. Introduction and Theoretical Development,* Cambridge University Press, Cambridge.
- Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, **122**, 108–124.
- Köhn, D., Kurzmann, A., Przebindowska, A., de Nil, D. & Bohlen, T., 2010. 2D elastic full waveform tomography of synthetic marine reflection seismic data, *DGG Mitt.*, **1/2010**, 23–27.
- Lévêque, J.J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, **115**, 313–318.
- Liu, D.C. & Nocedal, J., 1989. On the limited-memory BFGS method for large-scale optimisation, *Math. Program.*, **45**, 503–528.
- Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.*, 96, 2383–2397.
- Liu, Q. & Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.*, 174, 265–286.
- Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics*, 56, 645–653.
- Meier, U., Curtis, A. & Trampert, J., 2007a. Fully nonlinear inversion of fundamental mode surface waves for a global crustal model, *Geophys. Res. Lett.*, **34**, doi:10.1029/2007GL030989.
- Meier, U., Curtis, A. & Trampert, J., 2007b. Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.*, 169, 706–722.
- Meju, M.A., 2009. Regularised extremal bounds analysis (REBA): an approach to quantifying uncertainty in nonlinear geophysical inverse problems, *Geophys. Res. Lett.*, 36, doi:10:1029/2008GL036407.
- Meju, M.A. & Sakkas, V., 2007. Heterogeneous crust and upper mantle across southern Kenya and the relationship to surface deformation as inferred from magnetotelluric imaging, *J. geophys. Res.*, **112**, doi:10:1029/2005JB004028.
- Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, 52, 1211–1228.
- Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data, *Geophysics*, 53, 750–759.
- Mora, P., 1989. Inversion=migration+tomography, *Geophysics*, 54, 1575–1586.
- Nissen-Meyer, T., Dahlen, F.A. & Fournier, A., 2007. Spherical Earth Fréchet sensitivity kernels, *Geophys. J. Int.*, 168, 1051–1066.
- Nolet, G., 1990. Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs, *J. geophys. Res.*, **95**, 8499–8512.
- Nolet, G., 2008, *A Breviary of Seismic Tomography*. Cambridge University Press, Cambridge.
- Operto, S., Ravaut, C., Importa, L., Virieux, J., Herrero, A. & Dell'Aversana, P., 2004. Quantitative imaging of complex structures from dense wideaperture seismic data by multi-scale traveltime and waveform inversion: a case study, *Geophys. Prospect.*, 52, 625–651.
- Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, 167, 495–503.
- Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model, *Geophysics*, 64, 888–901.

(A1)

A2)

- Pratt, R., Shin, C. & Hicks, G., 1998. Gauss-Newton and full Newton methods in frequency domain seismic waveform inversion, *Geophys. J. Int.*, 133, 341–362.
- Resovsky, J. & Trampert, J., 2003. Using probabilistic seismic tomography to test mantle velocity-density relationships, *Earth planet. Sci. Lett.*, **215**, 121–134.
- Ritsema, J., vanHeijst, H. & Woodhouse, J.H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, 286, 1925–1928.
- Santosa, F. & Symes, W.W., 1988. Computation of the Hessian for leastsquares solutions of inverse problems of reflection seismology, *Inverse Probl.*, 4, 211–233.
- Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007a. Finite-frequency sensitivity of body waves to anisotropy based upon adjoint methods, *Geophys. J. Int.*, **171**, 368–389.
- Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007b. Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.*, **168**, 1153–1174.
- Smithyman, B., Pratt, R.G., Hayles, J. & Wittebolle, R., 2009. Detecting near-surface objects with seismic waveform tomography, *Geophysics*, 74, WCC119–WCC127.
- Tape, C., Liu, Q. & Tromp, J., 2007. Finite-frequency tomography using adjoint methods—Methodology and examples using membrane surface waves, *Geophys. J. Int.*, 168, 1105–1129.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science*, **325**, 988–992.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based upon spectral-element and adjoint methods, *Geophys. J. Int.*, 180, 433–462.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49, 1259–1266.

- Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, **51**, 1893–1903.
- Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure appl. Geophys.*, 128, 365–399.
- Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, 2nd edn. Society for Industrial and Applied Mathematics, Philadephia, PA.
- Tong, J., Dahlen, F.A., Nolet, G. & Marquering, H., 1998. Diffraction effects upon finite-frequency travel times: a simple 2-D example, *Geophys. Res. Lett.*, 25, 1983–1986.
- Trampert, J., Deschamps, F., Resovsky, J. & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science*, 306, 853–856.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, 160, 195–216.
- Wu, R.-S. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, 50, 582–595.
- Wu, R.-S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics*, 52, 11–25.
- Yoshizawa, K. & Kennett, B.L.N., 2005. Sensitivity kernels for finitefrequency surface waves, *Geophys. J. Int.*, 162, 910–926.
- Zhao, L., Jordan, T.H., Olsen, K.B. & Chen, P., 2005. Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, *Bull. seism. Soc. Am.*, 95, 2066–2080.
- Zhou, Y., 2009. Multimode surface wave sensitivity kernels in radially anisotropic Earth media, *Geophys. J. Int.*, 176, 865–888.
- Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.*, **158**, 142– 168.

APPENDIX A: COMPARISON WITH THE DISCRETE ADJOINT METHOD

The discrete adjoint method is a special case of the more general continuous adjoint method, applied to physical systems that are governed by algebraic equations. These systems can be either inherently discrete, or they can be the result of a discretization process. In this section, we establish close links between the continuous adjoint method as presented in Section 2 and the discrete adjoint method as it is commonly used for 2-D full waveform tomography in the frequency domain (e.g. Dessa *et al.* 2004; Bleibinhaus *et al.* 2007, 2009; Smithyman *et al.* 2009). Our development is slightly more general than the one presented by Pratt *et al.* (1998) and Pratt (1999) because it does not rely on a least-squares misfit functional.

A1 First derivatives

We start with the generic form of the space-discretized wave equation in the frequency-domain

$$-\omega^2 \mathbf{M} \bar{\mathbf{u}}(\omega) + \mathbf{K} \bar{\mathbf{u}}(\omega) = \bar{\mathbf{f}}(\omega),$$

where **M**, **K** and $\bar{\mathbf{u}}$ represent the mass matrix, the stiffness matrix and a discrete version of the seismic displacement field. The variables ω and $\bar{\mathbf{f}}$ are the angular frequency and the discrete external force density. Defining the impedance matrix $\mathbf{L}(\omega) = -\omega^2 \mathbf{M} + \mathbf{K}$, we can rewrite eq. (A1) in the form of a simple matrix-vector equation

$$L\bar{u} = f.$$
 (

The matrix **L** now contains all the structural information of the Earth model, that is, $\mathbf{L} = \mathbf{L}(\mathbf{m})$. Furthermore, we assume that the model space \mathfrak{M} is finite-dimensional, so that any model $\mathbf{m} \in \mathfrak{M}$ can be written in the form of a vector with $n < \infty$ components

$$\mathbf{m} = (m_1, m_2, \dots, m_n). \tag{A3}$$

We are interested in the partial derivatives of the data functional $\mathcal{X}(\mathbf{m}) = \mathcal{X}[\bar{\mathbf{u}}(\mathbf{m})]$ with respect to the model parameters m_i

$$\frac{\partial \mathcal{X}}{\partial m_i} = \nabla_u \mathcal{X} \frac{\partial \bar{\mathbf{u}}}{\partial m_i} \,. \tag{A4}$$

To eliminate $\partial \bar{\mathbf{u}} / \partial m_i$ from eq. (A4) we differentiate the discrete wave eq. (A2) with respect to m_i

$$\frac{\partial \mathbf{L}}{\partial m_i} \mathbf{\bar{u}} + \mathbf{L} \frac{\partial \mathbf{\bar{u}}}{\partial m_i} = \mathbf{0}, \quad \Rightarrow \quad \frac{\partial \mathbf{\bar{u}}}{\partial m_i} = -\mathbf{L}^{-1} \frac{\partial \mathbf{L}}{\partial m_i} \mathbf{\bar{u}}.$$
(A5)

© 2011 The Authors, *GJI*, **185**, 775–798 Geophysical Journal International © 2011 RAS

798 A. Fichtner and J. Trampert

The appearance of the inverse L^{-1} in eq. (A5) is purely symbolic. It does not have to be computed in practice. Substituting eq. (A5) into eq. (A4) yields

$$\frac{\partial \mathcal{X}}{\partial m_i} = -\nabla_u \mathcal{X} \left(\mathbf{L}^{-1} \cdot \frac{\partial \mathbf{L}}{\partial m_i} \right) \bar{\mathbf{u}} = -\bar{\mathbf{u}} \left[\frac{\partial \mathbf{L}^T}{\partial m_i} (\mathbf{L}^{-1})^T \right] \nabla_u \mathcal{X} \,. \tag{A6}$$

We now define the discrete adjoint wavefield $\bar{\boldsymbol{u}}^{\dagger}$ as the solution of the adjoint equation

$$\mathbf{L}^{T}\bar{\mathbf{u}}^{\dagger} = -\nabla_{\!u}\mathcal{X}\,.\tag{A7}$$

Clearly, eq. (A7) corresponds to the adjoint equation $\nabla_u \mathbf{L}^{\dagger} \mathbf{\bar{u}}^{\dagger} = -\nabla_u \chi^{\dagger}$ that we found in the continuous case (eq. 21). With the help of the discrete adjoint field $\mathbf{\bar{u}}^{\dagger}$ we can now obtain a simple expression for $\partial \mathcal{X} / \partial m_i$

$$\frac{\partial \mathcal{X}}{\partial m_i} = \bar{\mathbf{u}}^{\dagger} \frac{\partial \mathbf{L}}{\partial m_i} \bar{\mathbf{u}} \,. \tag{A8}$$

As in the continuous case, the computation of the partial derivatives of \mathcal{X} reduces to the solution of the adjoint eq. (A7) with an adjoint source determined by the data functional. The continuous counterpart of eq. (A8) is $\nabla_m \mathcal{X} \,\delta \mathbf{m} = \langle \bar{\mathbf{u}}^{\dagger} \nabla_m \mathbf{L} \,\delta \mathbf{m} \rangle$ (eq. 22).

A2 Second derivatives

Following the recipe from the previous section, we can derive an expression for the second derivative of \mathcal{X} in terms of the adjoint field $\mathbf{\bar{u}}^{\dagger}$, as defined in eq. (A7). Differentiating \mathcal{X} first with respect to m_i and then with respect to m_j gives

$$H_{ij} = \frac{\partial^2 \mathcal{X}}{\partial m_i \partial m_j} = \bar{H}_{ij} + \nabla_u \mathcal{X} \frac{\partial^2 \bar{\mathbf{u}}}{\partial m_i \partial m_j}, \tag{A9}$$

where the components of the approximate Hessian are defined as

$$\bar{H}_{ij} = \frac{\partial \bar{\mathbf{u}}}{\partial m_i} (\nabla_u \nabla_u \mathcal{X}) \frac{\partial \bar{\mathbf{u}}}{\partial m_j} \,. \tag{A10}$$

In contrast to the full Hessian, H_{ij} , the approximate Hessian merely involves first derivatives which makes its practical computation via the standard adjoint method comparatively straightforward. To eliminate the second derivative $\frac{\partial^2 \hat{\mathbf{u}}}{\partial m_i \partial m_j}$ from eq. (A9), we differentiate the forward problem (A2) twice

$$\frac{\partial^2 \mathbf{L}}{\partial m_i \partial m_j} \bar{\mathbf{u}} + \frac{\partial \mathbf{L}}{\partial m_i} \frac{\partial \bar{\mathbf{u}}}{\partial m_j} + \frac{\partial \mathbf{L}}{\partial m_j} \frac{\partial \bar{\mathbf{u}}}{\partial m_i} + \mathbf{L} \frac{\partial^2 \bar{\mathbf{u}}}{\partial m_i \partial m_j} = \mathbf{0}.$$
(A11)

The rearrangement of eq. (A11) provides an explicit expression for $\frac{\partial^2 \tilde{\mathbf{u}}}{\partial m_i \partial m_i}$,

$$\frac{\partial^2 \bar{\mathbf{u}}}{\partial m_i \partial m_j} = -\mathbf{L}^{-1} \left(\frac{\partial^2 \mathbf{L}}{\partial m_i \partial m_j} \bar{\mathbf{u}} + \frac{\partial \mathbf{L}}{\partial m_i} \frac{\partial \bar{\mathbf{u}}}{\partial m_j} + \frac{\partial \mathbf{L}}{\partial m_j} \frac{\partial \bar{\mathbf{u}}}{\partial m_i} \right), \tag{A12}$$

that we substitute into eq. (A9)

$$H_{ij} = \bar{H}_{ij} - \nabla_{u} \mathcal{X} \left[\mathbf{L}^{-1} \left(\frac{\partial^{2} \mathbf{L}}{\partial m_{i} \partial m_{j}} \bar{\mathbf{u}} + \frac{\partial \mathbf{L}}{\partial m_{i}} \frac{\partial \bar{\mathbf{u}}}{\partial m_{j}} + \frac{\partial \mathbf{L}}{\partial m_{j}} \frac{\partial \bar{\mathbf{u}}}{\partial m_{i}} \right) \right]$$

$$= \bar{H}_{ij} - \left[\left(\frac{\partial^{2} \mathbf{L}}{\partial m_{i} \partial m_{j}} \bar{\mathbf{u}} + \frac{\partial \mathbf{L}}{\partial m_{i}} \frac{\partial \bar{\mathbf{u}}}{\partial m_{j}} + \frac{\partial \mathbf{L}}{\partial m_{j}} \frac{\partial \bar{\mathbf{u}}}{\partial m_{i}} \right)^{T} (\mathbf{L}^{-1})^{T} \right] \nabla_{u} \mathcal{X}.$$
(A13)

Again defining the adjoint field $\bar{\boldsymbol{u}}^{\dagger}$ as the solution of

$$\mathbf{L}^{T}\bar{\mathbf{u}}^{\dagger} = -\nabla_{u}\mathcal{X},\tag{A14}$$

yields the desired formula for H_{ii} that is free of the explicit inverse of L and that does not contain second derivatives of $\bar{\mathbf{u}}$

$$H_{ij} = \bar{H}_{ij} + \bar{\mathbf{u}}^{\dagger} \left(\frac{\partial^2 \mathbf{L}}{\partial m_i \partial m_j} \bar{\mathbf{u}} + \frac{\partial \mathbf{L}}{\partial m_i} \frac{\partial \bar{\mathbf{u}}}{\partial m_j} + \frac{\partial \mathbf{L}}{\partial m_j} \frac{\partial \bar{\mathbf{u}}}{\partial m_i} \right).$$
(A15)

Eq. (A15) is the discrete analogue of the Hessian for time- and space-continuous problems (eq. 44).