Geophysical Journal International

Geophys. J. Int. (2011) 187, 1604-1624

Resolution analysis in full waveform inversion

Andreas Fichtner and Jeannot Trampert

Department of Earth Sciences, Utrecht University, Utrecht, the Netherlands. E-mail: andreas_fichtner@gmx.de

Accepted 2011 September 5. Received 2011 July 12; in original form 2011 May 12

SUMMARY

We propose a new method for the quantitative resolution analysis in full seismic waveform inversion that overcomes the limitations of classical synthetic inversions while being computationally more efficient and applicable to any misfit measure. The method rests on (1) the local quadratic approximation of the misfit functional in the vicinity of an optimal earth model, (2) the parametrization of the Hessian in terms of a parent function and its successive derivatives and (3) the computation of the space-dependent parameters via Fourier transforms of the Hessian, calculated with the help of adjoint techniques. In the simplest case of a Gaussian approximation, we can infer rigorously defined 3-D distributions of direction-dependent resolution lengths and the image distortion introduced by the tomographic method. We illustrate these concepts with a realistic full waveform inversion for upper-mantle structure beneath Europe. As a corollary to the method for resolution analysis, we propose several improvements to full waveform inversion techniques. These include a pre-conditioner for optimization schemes of the conjugate-gradient type, a new family of Newton-like methods, an approach to adaptive parametrization independent from ray theory and a strategy for objective functional design that aims at maximizing resolution. The computational requirements of our approach are less than for a typical synthetic inversion, but yield a much more complete picture of resolution and trade-offs. While the examples presented in this paper are rather specific, the underlying idea is very general. It allows for problem-dependent variations of the theme and for adaptations to exploration scenarios and other wave-equation-based tomography techniques that employ, for instance, georadar or microwave data.

Key words: Fourier analysis; Inverse theory; Seismic tomography; Computational seismology; Theoretical seismology; Wave propagation.

1 INTRODUCTION

Full waveform inversion is a tomographic technique that is based on numerical wave propagation through complex media combined with adjoint or scattering integral methods for the computation of Fréchet kernels. The accurate and complete solution of the seismic wave equation ensures that information from the full seismogram can be used for the purpose of improved tomographic models. Originally conceived in the late 1970s and early 1980s (Bamberger *et al.* 1977, 1982; Tarantola 1984), realistic applications have become feasible only recently. Full waveform inversion can now be used to solve local-scale engineering and exploration problems (e.g. Smithyman *et al.* 2009; Takam Takougang & Calvert 2011), to study crustal-scale deformation processes (e.g. Bleibinhaus *et al.* 2007; Chen *et al.* 2007; Tape *et al.* 2009), to reveal the detailed structure of the lower mantle (e.g. Konishi *et al.* 2009; Kawai & Geller 2010) or to refine continental-scale models for tectonic interpretations and improved tsunami warnings (e.g. Fichtner *et al.* 2010; Hingee *et al.* 2011). While the tomographic method itself has advanced substantially, an essential aspect of the inverse problem has been ignored almost completely, despite its obvious socio-economic relevance: the quantification of resolution and uncertainties.

1.1 Resolution analysis in full waveform inversion

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). This equivalence, however, holds only in the impractical case where the misfit χ is equal to the L_2 waveform difference. Furthermore, the rigorous analysis of diffraction tomography is restricted to homogeneous or layered acoustic media. The resulting resolution estimates are too optimistic for realistic applications that suffer from modelling errors and the sparsity of noisy data (Bleibinhaus *et al.* 2009).

Re-use of this article is permitted in accordance with the Terms and Conditions set out at http://wileyonlinelibrary.com/onlineopen#OnlineOpen_Terms

(4)

Resolution analysis in full waveform inversion is complicated by many factors: (1) The data depend non-linearly on the model, meaning that the well-established machinery of linear inverse theory is not applicable (Backus & Gilbert 1967; Tarantola 2005). (2) A direct consequence of non-linearity is the appearance of multiple local minima (e.g. Gauthier *et al.* 1986). These may be avoided with the help of various multiscale approaches (e.g. Bunks *et al.* 1995; Sirgue & Pratt 2004; Ravaut *et al.* 2004; Fichtner *et al.* 2009), also known as frequency-hopping in the microwave imaging literature (e.g. Chew & Lin 1995). The convergence of all currently used multiscale approaches is, however, purely empirical. (3) Contrary to most linearized tomographies, the sensitivity matrix is not computed explicitly in full waveform inversion for reasons of numerical efficiency. This prevents a local analysis based, for instance, on the computation of the resolution and covariance operators for large linear systems (Nolet *et al.* 1999; Boschi 2003). (4) The size of the model space and the costs of the forward problem solution prohibit the application of probabilistic approaches that account for non-linearity using Monte Carlo sampling (Sambridge & Mosegaard 2002; Tarantola 2005) or neural networks (Devilee *et al.* 1999; Meier *et al.* 2007a,b).

In the absence of a quantitative means to assess resolution, arguments concerning the reliability of full waveform inversion images are mostly restricted to synthetic inversions for specific input structures, on the visual inspection of the tomographic images or on the analysis of the data fit. Synthetic inversions are known to be potentially misleading even in linearized tomographies (Lévêque *et al.* 1993). Visual inspection is equally inadequate because the appearance of small-scale heterogeneities is too easily mistaken as an indicator of high resolution. Finally, a good fit between observed and synthetic waveforms merely proves that the tomographic system has been solved, but not necessarily resolved.

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 deficiency is the source of much scepticism as to whether it is really worth the effort.

Despite the difficulties introduced by the non-linearity of full waveform inversion combined with the computational costs of the forward problem solution, ample information about local resolution can be inferred from the quadratic approximation of the misfit functional χ .

$$\chi(\mathbf{m}) = \chi(\widetilde{\mathbf{m}}) + \frac{1}{2} \int_{G} \int_{G} [\mathbf{m}(\mathbf{x}) - \widetilde{\mathbf{m}}(\mathbf{x})]^{T} \mathbf{H}(\mathbf{x}, \mathbf{y}) [\mathbf{m}(\mathbf{y}) - \widetilde{\mathbf{m}}(\mathbf{y})] d^{3}\mathbf{x} d^{3}\mathbf{y},$$
(1)

where

$$\mathbf{m}(\mathbf{x}) = [m_1(\mathbf{x}), m_2(\mathbf{x}), \dots, m_N(\mathbf{x})]^T$$
(2)

represents an earth model composed of N physical quantities. The components of **m** may, for instance, be the P velocity α , the S velocity β and density ρ , that is, $(m_1, m_2, m_3)^T = (\alpha, \beta, \rho)^T$. The optimal earth model $\widetilde{\mathbf{m}}$ is characterized by a zero Fréchet derivative, meaning that

$$\nabla_m \chi(\widetilde{\mathbf{m}}) \,\delta \mathbf{m} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\chi(\widetilde{\mathbf{m}} + \epsilon \,\delta \mathbf{m}) - \chi(\widetilde{\mathbf{m}}) \right] = 0 \tag{3}$$

for all model perturbations $\delta \mathbf{m}$. The Hessian **H** is a symmetric and bilinear operator that acts on the perturbation $\delta \mathbf{m} = \mathbf{m} - \mathbf{\tilde{m}}$ via a double integral over the model volume *G*.

1.2 The role of the Hessian in resolution analysis

The importance of the Hessian in local resolution analysis arises directly from the second-order approximation (1) but also from its relations to the posterior covariance, extremal bounds analysis and point-spread functions (PSFs).

1.2.1 Inferences on resolution and trade-offs from the local approximation

Locally, that is, in the vicinity of the optimum $\tilde{\mathbf{m}}$, the Hessian describes the geometry of χ in terms of its curvature or convexity. In this sense, **H** provides the most direct measure of resolution and trade-offs as it describes the change of the misfit when $\tilde{\mathbf{m}}$ is slightly perturbed to $\tilde{\mathbf{m}} + \delta \mathbf{m}$. The diagonal element $H_{ii}(\mathbf{x}, \mathbf{x})$ defines the local resolution of the model parameter m_i at position \mathbf{x} . The off-diagonal elements $H_{ij}(\mathbf{x}, \mathbf{x})|_{i\neq j}$ measure the trade-offs between m_i and model parameters $m_j|_{j\neq i}$ at position \mathbf{x} , that is the extent to which the model parameters are dependent. Similarly, the off-diagonal elements $H_{ij}(\mathbf{x}, \mathbf{y})|_{\mathbf{x}\neq \mathbf{y}}$ encapsulate spatial dependencies between model parameters m_i and m_j at different positions \mathbf{x} and \mathbf{y} . Large off-diagonal elements imply that simultaneous perturbations of different parameters or in different regions can compensate each other, to leave the misfit χ nearly unchanged.

1.2.2 Relation of the Hessian to the posterior covariance

A further interpretation of the Hessian is related to Bayesian inference (e.g. Jaynes 2003; Tarantola 2005) where the available information on a model **m** is expressed in terms of a probability density $\sigma(\mathbf{m})$. In the specific case of a linear forward problem and Gaussian distributions describing prior knowledge and measurement errors, $\sigma(\mathbf{m})$ takes the form

$$\sigma(\mathbf{m}) = \text{const.} e^{-\chi_g(\mathbf{m})}$$

with the misfit functional

$$\chi_g(\mathbf{m}) = \frac{1}{2} \int_G \int_G [\mathbf{m}(\mathbf{x}) - \widetilde{\mathbf{m}}(\mathbf{x})]^T \mathbf{S}^{-1}(\mathbf{x}, \mathbf{y}) [\mathbf{m}(\mathbf{y}) - \widetilde{\mathbf{m}}(\mathbf{y})] \,\mathrm{d}^3 \mathbf{x} \,\mathrm{d}^3 \mathbf{y}$$
(5)

and the posterior covariance S. The comparison between (5) and the quadratic approximation (1) suggests the interpretation of H in terms of the inverse posterior covariance in a local probabilistic sense.

1.2.3 Extremal bounds analysis

In addition to being the carrier of covariance information, the Hessian **H** provides the extremal bounds within which the optimal model $\tilde{\mathbf{m}}$ can be perturbed without increasing the misfit $\chi(\tilde{\mathbf{m}})$ beyond a pre-defined limit $\chi(\tilde{\mathbf{m}}) + \delta\chi$, where $\delta\chi$ is usually related to the noise in the data (Meju & Sakkas 2007; Meju 2009). The model \mathbf{m}^{extr} that extremizes the integral of the model parameter m_i over a specific region $G_{\delta} \subset G$, that is, $\int_{G_{\delta}} m_i(\mathbf{x}) d^3\mathbf{x}$, while increasing the misfit to $\chi(\tilde{\mathbf{m}}) + \delta\chi$, is given by (Fichtner 2010)

$$m_{j}^{\text{extr}}(\mathbf{x}) = \widetilde{m}_{j}(\mathbf{x}) - \lambda \int_{G_{\delta}} H_{ij}^{-1}(\mathbf{x}, \mathbf{y}) \,\mathrm{d}^{3}\mathbf{y} \qquad \text{with} \qquad \lambda = \pm (2\delta\chi)^{1/2} \left[\int_{G_{\delta}} \int_{G_{\delta}} H_{ii}^{-1}(\mathbf{x}, \mathbf{y}) \,\mathrm{d}^{3}\mathbf{x} \,\mathrm{d}^{3}\mathbf{y} \right]^{-1/2} \tag{6}$$

and no summation over the repeated indices *ii*. Eq. (6) involves the inverse Hessian \mathbf{H}^{-1} , interpreted already in terms of the local posterior covariance. Extremal bounds analysis therefore attaches a deterministic and quantitative meaning to an originally probabilistic concept. Large variances H_{ii}^{-1} imply that large perturbations of parameter m_i within the region G_{δ} do not increase the misfit beyond the admissible bound $\delta \chi$, meaning that m_i is poorly constrained inside G_{δ} . The presence of non-zero covariances $H_{ij}^{-1}|_{j\neq i}$ indicates that joint perturbations of all model parameters compensate each other, to allow for even larger perturbations of the parameter m_i that we wish to extremize in the above-mentioned sense.

1.2.4 Point-spread functions

Finally, we can relate the Hessian to PSFs or spike tests that are commonly used as a diagnostic tool for resolution and trade-offs in linearized tomographic problems (e.g. Spakman 1991; Zielhuis & Nolet 1994; Yu *et al.* 2002; Fang *et al.* 2010). For this, we consider a special type of synthetic inversion where the initial model $\mathbf{m}^{(0)}(\mathbf{y})$ is nearly equal to the optimal model $\mathbf{\tilde{m}}(\mathbf{y})$. The only deficiency of $\mathbf{m}^{(0)}(\mathbf{y})$ is the absence of a point perturbation of parameter m_i that is point-localized at position \mathbf{x} , that is,

$$\mathbf{m}^{(0)}(\mathbf{y}) = \widetilde{\mathbf{m}}(\mathbf{y}) - \delta \mathbf{m}(\mathbf{y}) = \widetilde{\mathbf{m}}(\mathbf{y}) - \begin{bmatrix} \underbrace{0}_{\delta m_1}, \dots, \underbrace{0}_{\delta m_{i-1}}, \underbrace{\delta(\mathbf{y} - \mathbf{x})}_{\delta m_i}, \underbrace{0}_{\delta m_{i+1}}, \underbrace{0}_{\delta m_i} \end{bmatrix}^T .$$
(7)

Using the quadratic approximation (1) and the definition of the Fréchet derivative (3), we find that the *j*-component of the Fréchet derivative $\nabla_m \chi$ evaluated at the initial model $\mathbf{m}^{(0)}$, that is, $\nabla_m \chi_i^{(0)}(\mathbf{x})$, is given by

$$\nabla_{m} \chi_{j}^{(0)}(\mathbf{x}) = \int_{G} \left[m_{k}^{(0)}(\mathbf{x}') - \widetilde{m}_{k}(\mathbf{x}') \right] H_{kj}(\mathbf{x}', \mathbf{y}) d^{3}\mathbf{x}' = -\int_{G} \delta m_{k}(\mathbf{x}') H_{kj}(\mathbf{x}', \mathbf{y}) d^{3}\mathbf{x}' = -H_{ij}(\mathbf{x}, \mathbf{y}),$$
(8)

where the summation over repeated indices is implicitly assumed from hereon. The first iteration of a gradient-based optimization scheme would then update $\mathbf{m}^{(0)}(\mathbf{x})$ to an improved model $\mathbf{m}^{(1)} \approx \widetilde{\mathbf{m}}$, the components of which are given by

$$m_{j}^{(1)}(\mathbf{x}) = m_{j}^{(0)}(\mathbf{x}) - \gamma \,\nabla_{m} \,\chi_{j}^{(0)}(\mathbf{x}) = m_{j}^{(0)}(\mathbf{x}) + \gamma \,H_{ij}(\mathbf{x},\mathbf{y})\,.$$
⁽⁹⁾

The scalar γ is the step length that minimizes χ along the local direction of steepest descent, $-\nabla_m \chi^{(0)}(\mathbf{x})$. Eq. (9) reveals that the Hessian $\mathbf{H}(\mathbf{x}, \mathbf{y})$ represents our blurred perception of a point-localized perturbation at position \mathbf{y} in a linearized tomographic inversion. The effect of the off-diagonal elements $H_{ij}|_{i\neq j}$ is to introduce unwanted updates of model parameters $m_j|_{i\neq j}$ that have initially not been perturbed.

In the restricted sense of eq. (9), the Hessian $H_{ij}(\mathbf{x}, \mathbf{y})$ is the PSF, that is, the response of model parameter m_j to a linearized spike test with a point perturbation of m_i at position \mathbf{x} . A fully non-linear spike test based on gradient optimization with multiple iterations will generally lead to a sharper reconstruction of the input spike, so that $\mathbf{H}(\mathbf{x}, \mathbf{y})$ can be considered a conservative estimate of the non-linear PSF. Throughout the following developments, we use the term PSF in the linearized sense as a synonym for the Hessian because it offers an intuitive interpretation of $\mathbf{H}(\mathbf{x}, \mathbf{y})$. Although the significance of the Hessian in local resolution analysis is evident, the efficient computation of \mathbf{H} in time-domain modelling of seismic wave propagation remains challenging. The most efficient approach involves a modification of the well-known adjoint method (e.g. Tarantola 1988; Tromp *et al.* 2005; Fichtner *et al.* 2006; Liu & Tromp 2006; Plessix 2006; Chen 2011) that allows us to compute \mathbf{H} applied to a model perturbation $\delta \mathbf{m}$, that is,

$$\int_{G} \mathbf{H}(\mathbf{x}, \mathbf{y}) \,\delta \mathbf{m}(\mathbf{y}) \,\mathrm{d}^{3}\mathbf{y} \tag{10}$$

using two forward and two adjoint simulations, as described in Santosa & Symes (1988), Fichtner & Trampert (2011) and Appendix A.

A model perturbation $\delta \mathbf{m}$ samples the Hessian via the integral (10); and by sampling **H** with a suitable set of model perturbations, we can gather as much second-derivative information as needed for our purposes, though at the expense of potentially prohibitive computational requirements. It is therefore the purpose of this paper to develop a sampling strategy of the Hessian that operates with as few model perturbations as possible while leading to an approximation of **H** that is physically meaningful and interpretable.

1.3 Outlook

This paper is organized as follows: We start in Section 2 with a brief description of a full waveform inversion for upper-mantle structure beneath Europe that will serve as both motivation and testing ground for the subsequent developments. In Section 3, we approximate the Hessian by a position-dependent Gaussian, the parameters of which can be computed efficiently via the Fourier transform of H(x, y) for a small set of wavenumber vectors. The Gaussian approximation can be generalized with the help of Gram–Charlier expansions that express H(x, y) in terms of a parent function and its successive derivatives. Following the theoretical developments, we demonstrate in Section 4 how the Gaussian approximation of the Hessian can be used to infer the image distortion introduced by the tomographic method, as well as the distribution of direction-dependent resolution lengths. Section 5 provides an intuitive interpretation of the physics behind the Fourier transformed Hessian. This interpretation partly motivates several improvements to full waveform inversion techniques proposed in Section 6. These include a new family of Newton-like methods, a pre-conditioner for gradient methods, an approach to adaptive parametrization independent from ray theory and a criterion for the design of misfit functionals aiming at maximum resolution. Finally, in Appendices A and B, we review the computation of Hessian kernels and multidimensional Gram–Charlier expansions.

2 THE TESTING GROUND: A REALISTIC FULL WAVEFORM INVERSION

To illustrate the theoretical developments of the following sections with practical applications, we consider a long-wavelength full waveform inversion for European upper-mantle structure (Fig. 1). While this example is very specific, all subsequent developments are valid also for smaller scale applications, for instance, in the context of active source exploration and engineering problems.

As data we used 2200 mostly vertical-component seismograms from 40 events that occurred in Europe and western Asia between 2002 and 2010. To quantify the discrepancies between observed and synthetic waveforms, we measured time- and frequency-dependent phase misfits within a period band from 100 to 300 s (Fichtner *et al.* 2009). In addition to surface waves, we also included long-period body waves and unidentified phases. The numerical modelling was based on a spectral-element discretization of the seismic wave equation, as described in Fichtner & Igel (2008).

To ensure the rapid convergence of the iterative misfit minimization, we implemented a 3-D initial model. The initial crustal structure is a long-wavelength equivalent (Fichtner & Igel 2008) of the maximum likelihood model of Meier *et al.* (2007a,b)) who inverted surface wave dispersion for crustal thickness and the isotropic *S* velocity β . Within the continental crust, the isotropic *P* wave speed α , and density ρ , are scaled to β as $\alpha = 1.5399\beta + 840$ m s⁻¹, and $\rho = 0.2277\beta + 2016$ kg m⁻³(Meier *et al.* 2007a). Velocities and density within the oceanic crust are fixed to the values of crust 2.0 (Bassin *et al.* 2000). As 3-D initial mantle structure we use the isotropic *S* velocity variations of model S20RTS (Ritsema *et al.* 1999) added to an isotropic version of the 1-D model PREM (Dziewonski & Anderson 1981) where the 220 km discontinuity has been replaced by a linear gradient. The scaling to α is depth-dependent, as inferred from *P*, *PP*, *PPP* and *PKP* traveltimes (Ritsema & van Heijst 2002). The initial density structure in the mantle is radially symmetric.

During the inversion, we successively update both β and α , as well as ρ . After seven conjugate-gradient iterations, the total misfit reduction reached nearly 50 per cent, whereas it was below 2 per cent for the seventh iteration alone. This indicates that the final model, shown in Fig. 1(b), is indeed close to the optimum. Despite being computed from long-period data, the model clearly reveals prominent features previously imaged in both body and surface wave tomographic studies (e.g. Spakman 1991; Zielhuis & Nolet 1994; Boschi *et al.* 2004; Peter *et al.* 2008; Schivardi & Morelli 2011; Schäfer *et al.* 2011). These include the low velocities beneath the Iceland hotspot and the western Mediterranean Basin, the slow-to-fast transition from Phanerozoic central Europe to the Precambrian east European platform, and the high velocities beneath the Hellenic arc.

Having reached a nearly optimal model, we can compute the PSF, $\mathbf{H}(\mathbf{x}, \mathbf{y})$, in the sense of Section 1.2.4. The result for a point-localized *S* velocity perturbation at 180 km below northern Germany is shown in Fig. 2.



Figure 1. (a) Distribution of earthquakes (blue dots) and stations (red dots). Black lines represent the central segments of surface wave ray paths. (b) Horizontal slices through the *S* velocity model at 100 and 200 km depth.



Figure 2. Horizontal (top) and vertical (below) slices through different components of the PSF $\mathbf{H}(\mathbf{x}, \mathbf{y})$, for a fixed \mathbf{x} at 180 km depth below northern Germany. The free space variable is \mathbf{y} . (a) The $\beta\beta$ -component of the Hessian, $H_{\beta\beta}(\mathbf{x}, \mathbf{y})$, that corresponds to a point-localized *S* velocity perturbation at \mathbf{x} . Solid and dashed red lines indicate the position of \mathbf{x} . Note that the colour scale is chosen such that negative values are emphasized. (b) The same as to the left, but for the $\beta\alpha$ -component of the PSF, $H_{\beta\alpha}(\mathbf{x}, \mathbf{y})$. The amplitude of the horizontal slice at 180 km depth is multiplied by 20 to enhance visibility. The colour scale is again adapted to emphasise negative values. (c) Slices through the $\beta\rho$ -component of the Hessian, $H_{\beta\rho}(\mathbf{x}, \mathbf{y})$. Note the condensed colour scale and the appearance of strong negative values near the surface.

The spatial extent of the $\beta\beta$ -component $H_{\beta\beta}(\mathbf{x}, \mathbf{y})$, displayed in Fig. 2(a), describes the trade-offs between an *S* velocity perturbation located at position \mathbf{x} , and the *S* velocity in the surrounding model volume. With few exceptions, $H_{\beta\beta}(\mathbf{x}, \mathbf{y})$ is positive, meaning that negative perturbations in β compensate positive ones, and vice versa. The black and grey-shaded volume where trade-offs between *S* velocity perturbations are most significant, is roughly N-S oriented with a weak tail extending beneath Scandinavia and the North Atlantic. Interpreted in terms of a PSF, the Hessian component $H_{\beta\beta}(\mathbf{x}, \mathbf{y})$ is the *S* velocity structure recovered in a single-iteration synthetic inversion where the input was a perturbation in β localized at position \mathbf{x} . The width of the PSF loosely defines a resolution length within which *S* velocity perturbations cannot be constrained independently. We will return to the resolution length concept in Section 4.

Fig. 2(b) shows the $\beta\alpha$ -component of the Hessian, $H_{\beta\alpha}(\mathbf{x}, \mathbf{y})$, which represents the trade-offs between *P* and *S* velocity perturbations. While the horizontal slice qualitatively resembles $H_{\beta\beta}(\mathbf{x}, \mathbf{y})$, the vertical slices reveal that trade-offs between α and β are restricted to shallow depths of less than 50 km. This was to be expected because our data set is dominated by Rayleigh waves that are affected by *P* velocity structure only near the surface.

The $\beta \rho$ -component $H_{\beta\rho}(\mathbf{x}, \mathbf{y})$, visualized in Fig. 2(c), reveals a surprising complexity that is reminiscent of the strongly oscillatory density kernels for Rayleigh waves (e.g. Takeuchi & Saito 1972; Cara *et al.* 1984). The absolute values of $H_{\beta\rho}(\mathbf{x}, \mathbf{y})$ are generally small compared to those of the other components, meaning that the trade-offs between *S* velocity and density are hardly significant.

In general, the details of the PSFs depend on both the data and the position of the point perturbation relative to sources and receivers. However, from a series of numerical experiments, we conclude that several characteristics of the PSFs are position-independent. These include the comparatively small amplitude of $H_{\beta\rho}$, the restriction of $H_{\beta\alpha}$ to shallow depths and the roughly bell-shaped geometry of $H_{\beta\beta}$.

3 FOURIER-DOMAIN APPROXIMATION OF THE HESSIAN - I: THEORY

To fully quantify resolution and trade-offs, PSFs should ideally be computed for every model parameter and for every point \mathbf{x} within the volume of interest. Since the resulting computational costs would clearly be prohibitive, we propose an efficient scheme for the approximation of PSFs. This is based on a parametrization of the Hessian, followed by the estimation of the space-dependent parameters with the help of Fourier transforms.

3.1 Parametrization of the PSF

As briefly discussed in Section 1.2, a series of model perturbations $\delta \mathbf{m}^{(1)}, \delta \mathbf{m}^{(2)}, \ldots, \delta \mathbf{m}^{(n)}$ applied to **H** as in eq. (10) can be considered samplers of the Hessian. Our task is to approximate **H** from a small number of samples such that the result is physically meaningful and interpretable. This goal can be reached most efficiently by introducing a parametrization of the Hessian that incorporates prior knowledge on its general characteristics. Motivated, for instance, by the bell-shaped geometry of $H_{\beta\beta}$, shown in Fig. 2(a), we may parametrize the PSF for a specific point perturbation at position **x** by the **y**-dependent Gaussian

$$H_{\beta\beta}(\mathbf{x}, \mathbf{y}) \approx M^{(0)}(\mathbf{x}) \frac{\sqrt{\det \mathbf{C}(\mathbf{x})}}{(2\pi)^{3/2}} \exp\left[-\frac{1}{2} \left[\mathbf{y} - \mathbf{z}(\mathbf{x})\right]^T \mathbf{C}(\mathbf{x}) \left[\mathbf{y} - \mathbf{z}(\mathbf{x})\right]\right],\tag{11}$$

which is a common approximation for optical systems including microscopes and telescopes (e.g. Bendinelli *et al.* 1987; Trujillo *et al.* 2001; Zhang *et al.* 2007; Pankajakshan *et al.* 2000). The symmetric matrix $\mathbf{C}(\mathbf{x})$ describes the width and orientation of the PSF centred at $\mathbf{z}(\mathbf{x})$. Ideally, the diagonal elements of $\mathbf{C}(\mathbf{x})$ should be large, to minimize the spatial extent of the PSF. The difference $\mathbf{\Delta}(\mathbf{x}) = \mathbf{z}(\mathbf{x}) - \mathbf{x}$ describes the distortion introduced by the tomographic inversion that maps a point perturbation at \mathbf{x} into a blurred heterogeneity with a centre of mass at position $\mathbf{z}(\mathbf{x})$. Finally, the zeroth moment or weight $M^{(0)}(\mathbf{x}) \approx \int_G H_{\beta\beta}(\mathbf{x}, \mathbf{y}) d^3\mathbf{y}$ is equal to the integral over the PSF for a point perturbation at location \mathbf{x} . While being convenient, the parametrization of \mathbf{H} in terms of Gaussians is also restrictive, and its adequacy is clearly problem-dependent. More general parametrizations may be found within the framework of Gram–Charlier expansions, discussed in Section 3.2 and in Appendix B.

We have, at this point, transferred the problem of computing the PSF for every point in space to the determination of the space-dependent parameters in the approximation (11). An efficient approach to the computation of $M^{(0)}(\mathbf{x})$, $\mathbf{z}(\mathbf{x})$ and $\mathbf{C}(\mathbf{x})$ rests on the simple recognition that the application of \mathbf{H} to sinusoidal model perturbations allows us to compute the Fourier transform of the Hessian: Choosing, for instance,

$$\delta\beta^{(1)}(\mathbf{x}) = \cos(\mathbf{k}^T \mathbf{x}) \quad \text{and} \quad \delta\beta^{(2)}(\mathbf{x}) = \sin(\mathbf{k}^T \mathbf{x}),$$
(12)

with a fixed wavenumber vector **k**, yields the Fourier transform $H_{\beta\beta}(\mathbf{x}, \mathbf{k})$:

$$\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k}) = \int_{G} H_{\beta\beta}(\mathbf{x}, \mathbf{y}) \,\delta\beta^{(1)}(\mathbf{y}) \,\mathrm{d}^{3}\mathbf{y} + i \int_{G} H_{\beta\beta}(\mathbf{x}, \mathbf{y}) \,\delta\beta^{(2)}(\mathbf{y}) \,\mathrm{d}^{3}\mathbf{y}
= \int_{G} H_{\beta\beta}(\mathbf{x}, \mathbf{y}) \,\cos(\mathbf{k}^{T}\mathbf{y}) \,\mathrm{d}^{3}\mathbf{y} + i \int_{G} H_{\beta\beta}(\mathbf{x}, \mathbf{y}) \,\sin(\mathbf{k}^{T}\mathbf{y}) \,\mathrm{d}^{3}\mathbf{y}.$$
(13)

As mentioned previously, the integrals over the Hessian times a model perturbation can be computed efficiently with a variant of the adjoint method that requires two forward and two adjoint simulations [see Fichtner & Trampert (2011) and Appendix A].

Equating the right-hand side of (13) with the Fourier transform of the parametrized Hessian—the Gaussian from eq. (11), in our case—yields

$$\widetilde{H}_{\beta\beta}(\mathbf{x},\mathbf{k}) = M^{(0)} \exp\left(i\mathbf{k}^T \mathbf{z} - \frac{1}{2}\mathbf{k}^T \mathbf{C}^{-1}\mathbf{k}\right).$$
(14)

Eq. (14) can be transformed into the linear system

$$\operatorname{Re}\log\frac{H_{\beta\beta}(\mathbf{x},\mathbf{k})}{M^{(0)}} = -\frac{1}{2}\mathbf{k}^{T}\mathbf{C}^{-1}\mathbf{k},$$
(15)

Im log
$$\frac{\widetilde{H}_{\beta\beta}(\mathbf{x},\mathbf{k})}{M^{(0)}} = \mathbf{k}^T \mathbf{z}$$
. (16)

Given the zero-wavenumber component $M^{(0)}$, three non-zero and linearly independent wavenumber vectors are required to determine \mathbf{z} , as well as three linear combinations of the matrix elements C_{ij}^{-1} . To invert \mathbf{C}^{-1} the real parts of $H_{\beta\beta}$ for three additional wavenumber vectors are required. It follows that the Hessian must be applied to a total of 10 model perturbations: one for the zero-wavenumber component $M^{(0)}$, three plus three for the real and imaginary parts of $\tilde{H}_{\beta\beta}$ required for the computation of \mathbf{z} and three for the real parts of $\tilde{H}_{\beta\beta}$ needed to complete \mathbf{C}^{-1} . The resulting space-domain approximation of the PSF can then be used to infer position-dependent resolution and trade-offs, as illustrated in Section 4.

3.2 Generalization: Gram-Charlier approximations

Although the parametrization of the PSF in terms of space-dependent Gaussians will be sufficient for many practical purposes, it remains desirable to generalize the approach from Section 3.1 such that it allows for improved approximations. Gram–Charlier expansions, described in Appendix B, offer such a generalization in a natural way. In the following, we assume that the Fourier transform of $H_{ij}(\mathbf{x}, \mathbf{y})$ with respect to \mathbf{y} has been computed through the application of the Hessian to sinusoidal model perturbations, as illustrated in eq. (13). Omitting the subscripts in H_{ij} for notational convenience, the most general wavenumber-domain Gram–Charlier approximation of $\tilde{H}(\mathbf{x}, \mathbf{k})$ then takes the form

$$\widetilde{H}(\mathbf{x}, \mathbf{k}) = \exp\left[\left(H^{(0)} - G^{(0)}\right) + i\mathbf{k}^{T}\left(\mathbf{H}^{(1)} - \mathbf{G}^{(1)}\right) + \frac{i^{2}}{2!}\mathbf{k}^{T}\left(\mathbf{H}^{(2)} - \mathbf{G}^{(2)}\right)\mathbf{k} + \cdots\right]\widetilde{G}(\mathbf{k}),$$
(17)

© 2011 The Authors, *GJI*, **187**, 1604–1624

```
Geophysical Journal International © 2011 RAS
```

1610 A. Fichtner and J. Trampert

where \tilde{G} is the Fourier transform of a parent function G, and $\mathbf{H}^{(i)}(\mathbf{x})$ and $\mathbf{G}^{(i)}(\mathbf{x})$ are the position-dependent cumulants of H and G, also introduced in Appendix B. The first three cumulants of $H(\mathbf{x}, \mathbf{y})$ for a fixed value of \mathbf{x} are related to the moments

$$M_{ijk\dots}^{(n)} = \int_{\mathbb{R}^3} (y_i y_j y_k \dots) H(\mathbf{x}, \mathbf{y}) \,\mathrm{d}^3 \mathbf{y}$$
(18)

via

$$H^{(0)} = \log M^{(0)}, \quad H_i^{(1)} = \frac{M_i^{(1)}}{M^{(0)}}, \quad H_{ij}^{(2)} = \frac{1}{M^{(0)}} \left(M_{ij}^{(2)} - \frac{M_i^{(1)} M_j^{(1)}}{M^{(0)}} \right).$$
(19)

Similar relations hold for the cumulants and moments of the parent function G. The first and second cumulants, $\mathbf{H}^{(1)}$ and $\mathbf{H}^{(2)}$, are the centre of mass and the inverse covariance of $H(\mathbf{x}, \mathbf{y})|_{\mathbf{x}=\text{const.}}$, respectively.

To allow for a good approximation with few coefficients, the parent function is usually chosen to have the same general characteristics as the function that we wish to expand. Assuming, for instance, that $H(\mathbf{x}, \mathbf{y})|_{\mathbf{x}=\text{const.}}$ is roughly bell-shaped, as the PSF in Fig. 2(a), we can choose the space-domain parent function to be the normalized Gaussian

$$G(\mathbf{y}) = \frac{\sqrt{\det \mathbf{C}}}{(2\pi)^{3/2}} \exp\left[-\frac{1}{2} \left(\mathbf{y} - \mathbf{z}\right)^T \mathbf{C} \left(\mathbf{y} - \mathbf{z}\right)\right]$$
(20)

with the x-dependent cumulants

$$G^{(0)} = 0, \quad \mathbf{G}^{(1)} = \mathbf{z}, \quad \mathbf{G}^{(2)} = \mathbf{C}^{-1}, \quad \mathbf{G}^{(n>2)} = \mathbf{0}.$$
 (21)

Since the values of the first and second cumulants are still undetermined at this point of the development, we are free to impose $\mathbf{G}^{(1)} = \mathbf{H}^{(1)}$ and $\mathbf{G}^{(2)} = \mathbf{H}^{(2)}$, so that (17) reduces to the classical Gram–Charlier A series (e.g. Samuelson 1943; Cramér 1957)

$$\widetilde{H}(\mathbf{x}, \mathbf{k}) = M^{(0)} \exp\left[\frac{i^3}{3!} H^{(3)}_{ijk} k_i k_j k_k + \frac{i^4}{4!} H^{(4)}_{ijkl} k_i k_j k_k k_l + \dots\right] \exp\left(i\mathbf{k}^T \mathbf{z} - \frac{1}{2} \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k}\right).$$
(22)

Eq. (17) and its specialized form (22) are general expansions of the PSF in terms of its cumulants. The Gaussian parametrization from Section 3.1 is, in this sense, a first-order or low-wavenumber approximation. Higher order terms in the Gram–Charlier series involve higher cumulants and account for the shorter wavelength variations of the PSF.

Similar to the first-order approximation of eq. (14), the coefficients of a general Gram–Charlier expansions can be computed from the Fourier transforms for sufficiently many independent wavenumber vectors \mathbf{k} . The space-domain approximation corresponding to (22) is given by (see Appendix B)

$$H(\mathbf{x}, \mathbf{y}) = M^{(0)} \frac{\sqrt{\det \mathbf{C}}}{(2\pi)^{N/2}} \left[1 + \frac{(-1)^3}{3!} H^{(3)}_{ijk} \frac{\partial}{\partial y_i} \frac{\partial}{\partial y_j} \frac{\partial}{\partial y_k} + \frac{(-1)^4}{4!} H^{(4)}_{ijkl} \frac{\partial}{\partial y_i} \frac{\partial}{\partial y_j} \frac{\partial}{\partial y_k} \frac{\partial}{\partial y_l} + \cdots \right] \exp\left[-\frac{1}{2} (\mathbf{y} - \mathbf{z})^T \mathbf{C} (\mathbf{y} - \mathbf{z}) \right].$$
(23)

While the expansions (17) and (23) are very general, their physical interpretation is complicated by their complexity. We will therefore adhere to the low-wavenumber Gaussian approximation from Section 3.1 throughout the following examples.

4 FOURIER-DOMAIN APPROXIMATION OF THE HESSIAN - II: APPLICATION

To make the previous developments explicit, we return to the full waveform inversion example introduced in Section 2. As a preparatory step we subdivide the study region into nearly equal volume blocks, shown in Fig. 3. These introduce a coordinate system that is approximately cartesian over the expected width of the PSFs. The blocks, also used to parametrize the tomographic model from Fig. 1, are $1^{\circ} \times 1^{\circ}$ wide and 10 km deep. We then adopt the Gaussian parametrization of $H_{\beta\beta}$, proposed in eq. 11.



Figure 3. (Left) Subdivision of the model volume into blocks that are $1^{\circ} \times 1^{\circ}$ wide and 10 km deep. The blocks introduce a coordinate system that is nearly cartesian over the width of the PSFs. (Right) The blocks are used also to parametrize the tomographic model. This figure shows a small part of the *S* velocity model from Fig. 1 at 100 km depth. Superimposed are the local coordinate lines that define the boundaries of the blocks. Arrows indicate the local *x*- and *y*-directions.



Figure 4. (a) The zero-wavenumber perturbation $\delta\beta(\mathbf{k} = \mathbf{0})$ is equal to 1 m s⁻¹ throughout the model volume. (b) Horizontal slices at 100 km, 200 km and 300 km depth through the Fourier transform of the Hessian at zero wavenumber, $\widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k} = \mathbf{0})$.

4.1 The zero-wavenumber component: relative weight of PSFs

We obtain the Fourier transform at zero wavenumber, $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k} = \mathbf{0})$, from the application of $H_{\beta\beta}(\mathbf{x}, \mathbf{y})$ to an *S* velocity perturbation $\delta\beta(\mathbf{x})$ that is constant throughout the model volume, as shown in Fig. 4(a). The result is—irrespective of any parametrization or approximation—the space-dependent zeroth moment

$$M^{(0)}(\mathbf{x}) = \widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k} = \mathbf{0}) = \int_{G} H_{\beta\beta}(\mathbf{x}, \mathbf{y}) \,\mathrm{d}^{3}\mathbf{y} \,.$$
(24)

Horizontal slices though $M^{(0)}(\mathbf{x})$ at 100, 200 and 300 km depth are displayed in Fig. 4(b). As expected, the zeroth moment is mostly positive, with few exceptions at shallow depth confined to the immediate vicinity of sources or receivers. Clearly visible are diffuse versions of ray bundles that connect the epicentres along the periphery of the European plate to those regions with the highest station density. Furthermore, the zeroth moment decreases with depth below 200 km, as expected for a surface wave dominated data set.

According to eq. (24), the zeroth moment is equal to the integral over the PSF for a point perturbation at position \mathbf{x} . It follows that $M^{(0)}(\mathbf{x})$ is small when the misfit χ is nearly unaffected by a point perturbation at \mathbf{x} . In contrast, $M^{(0)}(\mathbf{x})$ is large when the PSF has a high amplitude, a large spatial extent, or both. In this sense, the spatial variability of $M^{(0)}(\mathbf{x})$ reflects the relative weight of neighbouring PSFs. Information on the resolution length is not contained in the zeroth moment.

From the perspective of gradient-based optimization, $M^{(0)}(\mathbf{x})$ is proportional to the direction of steepest descent evaluated at a model $\tilde{\mathbf{m}} + \delta \mathbf{m}$ that differs from the optimal model $\tilde{\mathbf{m}}$ by a constant *S* velocity perturbation. In the case of perfect data coverage, we would therefore expect $M^{(0)}(\mathbf{x})$ to be constant throughout the model volume. The obvious departures from this ideal scenario, seen in Fig. 4(b), result from the limited sensitivity to deep structure in a surface wave dominated tomography, and from the clustering of sources and receivers. These clusters dominate the misfit, which then leads to a heterogeneous direction of steepest descent.

4.2 Real and imaginary parts of the Fourier transform: image distortion

Tomographic inversions map point perturbations into blurred versions of themselves. The centre of mass of the blurred image—that is the position where the heterogeneity is visually perceived—does not generally coincide with the location of the original point perturbation. This distortion can be quantified, at least in a linearized sense, using the approach described in Section 3.

Based on eq. (12), we compute the spectrum $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$ for a non-zero wavenumber \mathbf{k} by applying the Hessian $H_{\beta\beta}(\mathbf{x}, \mathbf{y})$ to the sinusoidal model perturbations $\delta\beta(\mathbf{x}) = \cos(\mathbf{k}^T \mathbf{x})$ and $\delta\beta(\mathbf{x}) = \sin(\mathbf{k}^T \mathbf{x})$. For our example, we choose the wavelength of the oscillations to be 2000 km. The model perturbations together with the real and imaginary parts of $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$ are shown in Fig. 5 for a wavenumber vector pointing in the *x*-direction of the coordinate system introduced in Fig. 3 (nearly N–S in geographic coordinates).

Both the real and imaginary parts reveal a similar oscillatory pattern as the sinusoidal model perturbations from which they originate. This behaviour is to be expected provided that any PSF is roughly a space-shifted version of the PSF for a perturbation at the coordinate origin $\mathbf{x} = \mathbf{0}$, that is, $H_{\beta\beta}(\mathbf{x}, \mathbf{y}) \approx H_{\beta\beta}(\mathbf{0}, \mathbf{y} - \mathbf{x})$. Deviations from exact sinusoidal oscillations therefore reflect the spatially variable shape and amplitude of the PSFs.

When interpreted within the framework of the Gaussian approximation (eqs 11 and 14), we can use the spectrum from Fig. 5 to compute the distortion

$$\Delta(\mathbf{x}) = \mathbf{z}(\mathbf{x}) - \mathbf{x} \tag{25}$$

which is the distance between a point-localized perturbation at position \mathbf{x} , and the centre of mass of the corresponding PSF, $\mathbf{z}(\mathbf{x})$. The distortion measures the displacement of a point perturbation caused by the tomographic inversion. Upon inserting the definition (25) into eq. (14), we



Figure 5. Horizontal slices through the sinusoidal model perturbations and the corresponding real and imaginary parts of $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$. The black curves in the spectral maps delineate the zero lines in the model perturbations and are intended to guide the eye.



Figure 6. (a) The lateral component of the distortion $\Delta(\mathbf{x})$ at 100 and 200 km depth, respectively. The length of the red line in the figure headings is the scale for the length of the arrows. (b) Surface projections of the central segments of the ray paths connecting sources (blue dots) to receivers (red dots).

obtain

$$\widetilde{H}_{\beta\beta}(\mathbf{x},\mathbf{k})\exp\left(-i\mathbf{k}^{T}\mathbf{x}\right) = M^{(0)}\exp\left(i\mathbf{k}^{T}\boldsymbol{\Delta}\right)\exp\left(-\frac{1}{2}\mathbf{k}^{T}\mathbf{C}^{-1}\mathbf{k}\right).$$
(26)

The three components of the distortion $\mathbf{\Delta}$ can then be obtained from the evaluation of eq. (26) for three linearly independent wavenumber vectors. We note that the calculation of the complex argument $\mathbf{k}^T \mathbf{\Delta}$ needs regularized because phase jumps appear where $|\widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})|$ is small. This regularization can be achieved by adding a constant $\epsilon > 0$ to the real part of (26), so that $\mathbf{k}^T \mathbf{\Delta}$ is forced to zero as $|\widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})|$ drops below ϵ .

The lateral distortion, that is, the horizontal component of Δ , computed from the spectra for two linearly independent wavenumber vectors, is shown in Fig. 6. The interpretation must account for the regularization, which in this particular case means that Δ is strongly attenuated in regions where the absolute value of the spectrum shown in Fig. 5 drops below 0.5×10^{-15} rad s m⁻⁴. Loosely speaking, the distortion is meaningful only in regions where the PSFs are significantly non-zero.

Throughout central Europe, the lateral distortion is below 300 km, which is due to the high station density. Large distortions appear to be rather localized, for instance, east of the Caspian Sea where ray paths are almost exclusively N–S oriented. As intuitively expected, Δ slightly increases with increasing depth.

4.3 Resolution lengths

One of the most relevant quantities to be considered in resolution analysis is the resolution length, that is, the distance within which two point perturbations cannot be constrained independently.

The resolution length is the direction-dependent extent of the PSF that defines a region where the original point perturbation trades off significantly with neighbouring heterogeneities. Within the Gaussian approximation (11), the width of a PSF is controlled by the positive definite matrix C(x), the inverse of which can be estimated from the spectrum (14).

We use the standard deviation of the Gaussian in a direction specified by the unit vector e,

$$\ell_{\mathbf{e}} = \frac{1}{\sqrt{\mathbf{e}^T \mathbf{C} \, \mathbf{e}}}\,,\tag{27}$$



resolution length [km]

Figure 7. Horizontal slices at 100 and 200 km depth through the resolution length distributions in the directions indicated by the arrows. (a) *x*-direction of the local coordinate system from Fig. 3 (roughly N–S in central and western Europe). (b) *y*-direction of the local coordinate system from Fig. 3 (roughly E–W in central and western Europe).

as our definition of the direction-dependent resolution length. This means in the context of our specific example that *S* velocity perturbations separated by less than the resolution length effectively appear as one heterogeneity, instead of having clearly separate identities.

Distributions of horizontal resolution lengths in two perpendicular directions are shown in Fig. 7. The resolution length in x-direction (nearly N–S in central and western Europe) reaches its minimum of 300 km beneath the Balkan peninsula, in agreement with the dense clustering of E–W-oriented ray paths in that region. Additional minima can be observed, for instance, in the vicinity of events along the Mid-Atlantic ridge from where waves travel westwards to stations in central and southern Europe. The resolution length in the local y-direction (nearly E–W in central and western Europe) is longer than in x-direction beneath the Balkans but significantly shorter beneath the North Atlantic. This, again, is in accord with the predominance of N–S-oriented ray paths that connect sources along the northern Mid-Atlantic ridge to stations on the continent. Throughout most of Europe, the resolution length in any direction varies between 400 and 600 km, which is close to the wavelength of the 120 s surface waves in our data set.

5 THE PHYSICS BEHIND: THE RETURN OF SYNTHETIC INVERSIONS?

In the following paragraphs, we complement the formal treatment of the Fourier transform by an intuitive interpretation. This is intended to shed light onto the physics behind the oscillatory behaviour of the Fourier transformed Hessian, and to reveal a link between the width of Fresnel zones and the resolution length. To facilitate our arguments, we restrict ourselves to the scalar case with one model parameter only.

Our interpretation rests on the second-order approximation of the misfit χ (eq. 1), based upon which the Fréchet derivative $\nabla_m \chi$, evaluated at an initial model $m^{(0)} = \tilde{m} + \delta m$, can be expressed in terms of the Hessian.

$$\nabla_m \chi(\mathbf{x}) = \int_G H(\mathbf{x}, \mathbf{y}) \,\delta m(\mathbf{y}) \,\mathrm{d}^3 \mathbf{y} \,. \tag{28}$$

For a sinusoidal perturbation $\delta m(\mathbf{x}) = \sin(\mathbf{k}^T \mathbf{x})$, the Fréchet derivative (28) equals the Fourier transformed Hessian, that is, $\nabla_m \chi(\mathbf{x}) = \hat{H}(\mathbf{x}, \mathbf{k})$. Since $\nabla_m \chi(\mathbf{x})$ is the local direction of steepest ascent, the first iteration of a conjugate-gradient-type synthetic inversion updates $m^{(0)}$ to

$$m^{(1)}(\mathbf{x}) = m^{(0)}(\mathbf{x}) - \gamma \,\widetilde{H}(\mathbf{x}, \mathbf{k}), \tag{29}$$

where γ is the optimal step length. Thus, in the ideal but unrealistic case of perfect convergence within one iteration, the spectrum $\tilde{H}(\mathbf{x}, \mathbf{k})$ should exactly reproduce the oscillations of $\delta m(\mathbf{x})$. The practically observed differences between $m^{(1)}(\mathbf{x})$ and the optimum $\tilde{m}(\mathbf{x})$ result from the smearing of $\delta m(\mathbf{x})$ by the PSF $H(\mathbf{x}, \mathbf{y})$ via the integral in eq. (28). The nature of the smearing operation depends on the characteristics of the PSFs, that are governed by the source–receiver geometry and the waveforms included in the misfit χ .

For further illustrations of the smearing procedure, we consider the simplistic configuration shown in Fig. 8(a). The sources and receivers are oriented horizontally, parallel to the zero lines of δm . Each receiver records one isolated *S* wave that propagates from one source with the *S* velocity β .

We can infer the geometry of the PSFs for this configuration from plausibility arguments: The PSF $H(\mathbf{x}, \mathbf{y})$ is the response of the tomographic system to a point perturbation at position \mathbf{x} . While this has to be understood in the linearized sense (see Section 1.2.4), the Hessian nevertheless accounts for both first- and second-order scattering. When \mathbf{x} is located inside the first Fresnel zone, the corresponding PSF is non-zero, because first- and second-order scattering affect the measurement. The shape of the PSF roughly follows the first Fresnel zone, which is the region where different perturbations cannot be independently identified. In particular, the maximum width d of the PSF is close to the maximum width of the first Fresnel zone, that is, $d \approx \sqrt{\beta TL}$, where T and L are the dominant period and the epicentral distance, respectively. When the position of the point perturbation, \mathbf{x} , is outside the first Fresnel zone, both first- and second-order scattering PSF is then effectively zero. In conclusion, the PSFs for our simplistic setup occupy nearly cigar-shaped regions connecting sources and receivers.



Figure 8. Schematic illustrations of the Hessian applied to a sinusoidal model perturbation $\delta m(\mathbf{x})$, for different source–receiver geometries. (a) The PSFs for horizontally oriented sources (•) and receivers (+) roughly occupy the blue-shaded cigar-shaped regions aligned parallel to the oscillations of $\delta m(\mathbf{x})$. The maximum width *d* of the PSF in this configuration is approximately equal to the width of the first Fresnel zone, that is, $d \approx \sqrt{\beta TL}$, with the source–receiver distance *L*. The resulting integral $\int_G H(\mathbf{x}, \mathbf{y}) \delta m(\mathbf{y}) d^3 \mathbf{y}$, shown below, is a blurred but still recognizable version of $\delta m(\mathbf{x})$. (b) The same as to the left, but for a vertically oriented source–receiver configuration. The PSFs are perpendicular to the sinusoidal variations of $\delta m(\mathbf{x})$. Positive and negative contributions tend to cancel, leading to a nearly zero integral $\int_G H(\mathbf{x}, \mathbf{y}) \delta m(\mathbf{y}) d^3 \mathbf{y}$. (c) The same as in the left column, but for a larger period *T*. As the dominant period increases, the PSFs broaden, thus leading to more heavily blurred and attenuated integral $\int_G H(\mathbf{x}, \mathbf{y}) \delta m(\mathbf{y}) d^3 \mathbf{y}$.

In our first conceptual example from Fig. 8(a), the dominant period *T* is sufficiently small to generate PSFs that fit within less than half an oscillation period of δm . The integral in eq. (28) therefore reproduces a slightly blurred but still clearly recognizable version of the sinusoidal δm . Seen from the data perspective, waves propagating perpendicular to the oscillations are continuously distorted, thus increasing the misfit from the minimum $\chi(\tilde{m})$ to $\chi(\tilde{m} + \delta m)$.

For a vertically oriented source–receiver configuration, schematically illustrated in Fig. 8(b), the PSFs are parallel to the wavenumber vector. Positive and negative contributions to the integral (28) nearly cancel, which then leads to a Fréchet derivative (28) that is close to zero. Also the misfit is hardly affected, because the wavefield distortions accumulated within the positive and negative parts of δm compensate each other.

The previous line of arguments allows us to interpret the spectrum $\tilde{H}(\mathbf{x}, \mathbf{k})$ loosely as a measure for the information transmitted across the model in the directions perpendicular to the wavenumber vector \mathbf{k} . Large absolute values of $\tilde{H}(\mathbf{x}, \mathbf{k})$ correspond, in this sense, to a large amount of information transported by waves propagating nearly perpendicular to \mathbf{k} .

From the amplitudes of the oscillations in $\tilde{H}(\mathbf{x}, \mathbf{k})$, we can infer the resolution length. In fact, based on the spectrum of the Gaussian (14) and the definition of the resolution length $\ell_{\mathbf{e}}$ (27), we find that $|\tilde{H}(\mathbf{x}, \mathbf{k})|$ is approximately proportional to $\exp(-|\mathbf{k}|^2 \ell_{\mathbf{e}}^2/2)$. Thus, as the PSF or the first Fresnel zone extends with increasing period *T*, the smoothing integral (28) reduces the amplitude of the spectrum $\tilde{H}(\mathbf{x}, \mathbf{k})$, as shown in Fig. 8(c). The amplitude reduction then corresponds to an increasing resolution length $\ell_{\mathbf{e}}$.

An immediate consequence of this relation is the variation of the resolution length along the ray path. It attains a maximum near the centre of the ray path, and minima in the vicinity of sources and receivers where the lateral extent of the Fresnel zone or PSF is minimal.

It is a curious aspect of the previous interpretation that we can obtain information on the Hessian, and thus on resolution and trade-offs, from a series of linearized synthetic inversions in the sense of eqs (28) and (29). This information can be made as complete as needed by replacing the Gaussian approximation (11) by Gram–Charlier expansions (Section 3.2) or any other suitable parametrization of the Hessian. The required synthetic inversions differ from the common resolution tests of the chequer board type in that they start from an initial model $\mathbf{m}^{(0)}$ that is equal to an already computed optimal model $\mathbf{\tilde{m}}$ plus a sinusoidal perturbation $\delta \mathbf{m}$.

While synthetic inversions with one single input model are known to be incomplete and potentially misleading (Lévêque *et al.* 1993), the Fourier-domain approximation of the Hessian builds on a small set of linearly independent oscillatory input models that sample the PSFs in different directions. Synthetic inversions are, from this perspective, indeed a useful tool, but they need to be combined for a meaningful resolution analysis.

6 POTENTIALS AND PERSPECTIVES

The approximation of the Hessian with the help of parametrizations such as those suggested in Section 3 does not rely on the proximity of an optimal model. It may, in fact, be used at any stage of an iterative inversion. Thus, as a corollary to the previous developments, we can propose several improvements to full waveform inversion techniques. These include a new class of Newton-like methods, a pre-conditioner

for optimization schemes of the conjugate-gradient type, an approach to adaptive parametrization and a strategy for objective functional design. This section is primarily intended to be thought-provoking. All of the proposed improvements require further research that is beyond the scope of this work.

6.1 Non-linear optimization

The feasibility of full waveform inversion relies on the efficiency of the non-linear optimization schemes used to minimize the misfit χ . 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), variable-metric methods (e.g. Liu & Nocedal 1989; Brossier *et al.* 2009) and Newton-like schemes including the Gauss–Newton and Levenberg–Marquardt methods. (e.g. Pratt *et al.* 1998; Epanomeritakis *et al.* 2008).

6.1.1 Pre-conditioning of descent directions

While the efficiency of Newton's method and its variants in large-scale full waveform inversion is still debated, conjugate-gradient schemes enjoy widespread popularity. In their pure form, however, conjugate-gradient algorithms may converge slowly because Fréchet kernels are small in weakly covered areas and extremely large near sources and receivers. The iterative inversion therefore tends to generate excessive heterogeneities around hypocentres and station arrays, unless the descent directions are pre-conditioned. The pre-conditioner—ideally close to the inverse Hessian—acts to balance the extreme and the weak contributions, thus leading to faster convergence and to physically more reasonable updates.

Owing to the difficulties involved in the computation of the inverse Hessian, several intuitive pre-conditioning strategies are commonly applied. These include smoothing and the correction for the geometric spreading of the forward and adjoint fields (Igel *et al.* 1996; Fichtner *et al.* 2009, 2010).

The Fourier transform of the Hessian $\tilde{H}(\mathbf{x}, \mathbf{k})$, interpreted in terms of an ascent direction in Section 5, provides a pre-conditioner that can be computed efficiently. In the hypothetical case of perfect coverage, $\tilde{H}(\mathbf{x}, \mathbf{k})$ is expected to be a scaled version of the oscillatory model perturbations used for its computation. The synthetic inversion from eq. (29) would then converge in one iteration. In a realistic case with irregular coverage, however, the first update $m^{(1)}$ and the target model $\tilde{m} + \delta m$ are not identical, as illustrated in Fig. 5.

Seen from the perspective of the Gaussian approximation (14), the deviations of $\tilde{H}(\mathbf{x}, \mathbf{k})$ from a harmonic function result from the spatially variable zeroth moment $M^{(0)}(\mathbf{x})$, the non-zero distortion $\Delta(\mathbf{x}) = \mathbf{z}(\mathbf{x}) - \mathbf{x}$, and the variable width of the PSFs encapsulated in the symmetric matrix $\mathbf{C}(\mathbf{x})$. While corrections of the ascent direction $\tilde{H}(\mathbf{x}, \mathbf{k})$ for the distortion and the width of the PSF are wavenumber-specific, a correction for the zeroth moment is generally possible for any initial model.

The proposed pre-conditioning consists in the division of the pure ascent direction by $M^{(0)}(\mathbf{x})$, computed for the current model. This operation, illustrated in Fig. 9, generally requires regularization to avoid division by zero. Fig. 9(a) shows the oscillatory model perturbation used to compute the real part of $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$ in Fig. 9(b). The amplitudes of $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$ are weak in less covered regions, and large where sources and receivers cluster. The division of $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$ (Fig. 9d) balances the amplitudes, which leads to a pre-conditioned ascent direction that reproduces the oscillations of the input model much more closely than $\tilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$. The computationally inexpensive division by the zeroth moment can thus be expected to result in a faster convergence of gradient-based inversions.



Figure 9. (a) Horizontal slice through the sinusoidal model perturbation $\delta\beta(\mathbf{x}) = \cos(\mathbf{k}^T \mathbf{x})$, used to generate the real part of $\widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$, shown in (b). The black curves correspond to the zero lines of $\delta\beta$. (c) The zero-wavenumber component of the Hessian spectrum, $\widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k} = \mathbf{0})$. (d) The pre-conditioned descent direction $\widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k}) / \widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k} = \mathbf{0})$, which reproduces the oscillatory behaviour of $\delta\beta$ more closely than the pure descent direction $\widetilde{H}_{\beta\beta}(\mathbf{x}, \mathbf{k})$.

1616 A. Fichtner and J. Trampert

6.1.2 A new class of Newton-like methods

While the convergence rate of Newton-like methods can be nearly quadratic, their applicability is severely limited by the need to solve Newton's equation for the descent direction. Since the (approximate) Hessian can neither be computed nor stored explicitly for realistic problems, Newton's equation is solved iteratively using, for instance, conjugate-gradient algorithms, where each iteration requires the evaluation of the (approximate) Hessian times a model perturbation. The resulting computational costs have so far prevented the application of Newton-like methods in large-scale waveform inversions.

Parametrized versions of the Hessian, such as the ones proposed in Section 3, may offer a computationally more efficient alternative to conventional Newton-like schemes, because an approximate Hessian can be computed explicitly. The application of the parametrized Hessian to a descent direction—needed for the conjugate-gradient solution of Newton's equation—merely involves an integral over space, which greatly accelerates the computations. Furthermore, the storage requirements for the parametrized Hessian are negligible compared to those involved in the computation of Fréchet and Hessian kernels via adjoint techniques.

6.2 Adaptive parametrization

The uneven sampling of the Earth by seismic observables can lead to severe imaging artefacts unless the model parametrization is adapted to the spatially variable resolution. This issue is most critical when local and global data are combined. Approaches to the design of adaptive parametrizations mostly based on ray theory have been developed by various authors (e.g. Widiyantoro & van der Hilst 1997; Bijwaard *et al.* 1998; Boschi *et al.* 2004; Debayle & Sambridge 2004; Nolet & Montelli 2005; Bodin & Sambridge 2009; Schäfer *et al.* 2011). In general, the size of the basis functions is chosen inversely proportional to the density and azimuthal coverage of rays that correspond to well-defined seismic phases.

The problem arising in full waveform inversion is that the strongly application-specific misfit measures are applicable to complete seismograms, without being necessarily related to isolated phases for which seismic rays can be computed (e.g. Crase *et al.* 1990; Fichtner *et al.* 2008; Brossier *et al.* 2010; van Leeuwen & Mulder 2010; Bozdağ *et al.* 2011). The resolution length concept, introduced in Section 4.3, offers a natural solution to this problem. Instead of using the ray coverage as a proxy for resolution, we propose to choose the shape of the basis functions according to the resolution length in different directions.

The design of an adaptive parametrization can be incorporated into a multiscale approach where the dominant period decreases with increasing number of iterations (e.g. Bunks *et al.* 1995; Sirgue & Pratt 2004; Fichtner 2010): First, a very coarse parametrization is chosen *ad hoc* for the first iterations with the longest period data. When the model is sufficiently close to the optimum, the resolution lengths are computed, and then used to refine the parametrization. The iterative inversion is then continued with shorter period data until the next optimum is reached. This refinement/optimization procedure can then be repeated as often as needed.

6.3 Objective functional design

A further potential application of the previously developed methodology is the design of objective functionals, commonly constructed to meet criteria in the data space, such as robustness (Crase *et al.* 1990; Brossier *et al.* 2010) and separation of phase and amplitude (Luo & Schuster 1991; Gee & Jordan 1992; Fichtner *et al.* 2008; Bozdağ *et al.* 2011). However, the design of objective functionals with the aim to maximize resolution has so far received little attention in full waveform inversion (Maurer *et al.* 2009)—contrary to common practice in linear inverse problems where Backus–Gilbert theory provides all necessary tools (Backus & Gilbert 1967). This is mostly because a quantitative resolution analysis in full waveform inversion was not available. The resolution length computed from the Hessian may serve as a natural measure for the relative performance of objective functionals that already satisfy principal desiderata in the data space. It may as such also be used to assess the resolution capabilities of novel seismic observables, including rotations and strain (Bernauer *et al.* 2009; Ferreira & Igel 2009).

7 DISCUSSION

7.1 Computational requirements

Our approach to resolution analysis is based on Fourier transforms of the Hessian for a set of linearly independent wavenumber vectors. Accepting the Gaussian parametrization of $H(\mathbf{x}, \mathbf{y})$, as proposed in Section 3.1, 10 applications of the Hessian to sinusoidal model perturbations are needed to uniquely determine all space-dependent parameters, that is, $M^{(0)}(\mathbf{x})$, $\mathbf{z}(\mathbf{x})$ and $\mathbf{C}(\mathbf{x})$.

As described in Appendix A, each application of the Hessian to a model perturbation requires four wave field simulations: one for the regular forward field \mathbf{u} , one for the regular adjoint field \mathbf{u}^{\dagger} , one for the scattered forward field $\delta \mathbf{u}$ and finally one for the scattered adjoint field $\delta \mathbf{u}^{\dagger}$. The simulations of the regular forward and adjoint fields can be reused for every application of the Hessian to any model perturbation. Only the scattered forward and adjoint fields are perturbation-specific. This brings the total number of simulations to 22.

It is instructive to compare 22 to the number of simulations required for a typical synthetic inversion based on a conjugate-gradient method. Each iteration requires one forward and one adjoint simulation, plus two additional forward simulations to determine the optimal step length. It follows that 24 simulations allow us to perform six conjugate-gradient iterations in a synthetic inversion.

In realistic applications, however, the number of iterations is on the order of 20 (e.g. Tape *et al.* 2010; Fichtner *et al.* 2010), meaning that a synthetic inversion with only six iterations would be of little use. We therefore conclude that our approach is computationally more efficient than synthetic inversions, while providing a much more complete characterization of resolution.

7.2 Gram-Charlier expansions

A weak point in our development lies in the parametrization of the Hessian by a Gram–Charlier series—truncated, in the simplest case, after the first-order term. As shown in Appendix B, Gram–Charlier series can be considered a variant of Taylor series. It follows that these two expansions potentially share some of their disadvantageous properties: slow convergence or even divergence, depending on the function that we wish to expand.

The efficiency of Gram–Charlier series generally depends on the choice of the parent function. This choice was straightforward in the case of the $\beta\beta$ -component of the Hessian, $H_{\beta\beta}$, which is always roughly bell-shaped, that is, well-approximated by a Gaussian (see Fig. 2a). However, the off-diagonal elements $H_{\beta\alpha}$ and $H_{\rho\alpha}$, shown in Figs 2(b) and 2(c), are less predictable. This makes the choice of a suitable and generic parent function difficult or even impossible. As a result of this complication, we did not attempt to parametrize $H_{\beta\alpha}$ and $H_{\rho\alpha}$, meaning that we are currently not able to study interparameter trade-offs in a comprehensive way. Clearly, a more powerful parametrization of the Hessian would be a substantial improvement to our method.

7.3 Relation to migration deconvolution

Full waveform inversion can be regarded similar to various forms of migration (e.g. Tarantola 1984; Chavent & Plessix 1999; Nemeth *et al.* 1999). Reverse-time migration, in particular, can be interpreted as the first iteration of a full waveform inversion, provided that measurements are made only on body waves with an L_n norm misfit. A close link therefore exists between our approach and migration deconvolution used in exploration. Schuster & Hu (2000) computed migration PSFs for homogeneous media analytically. In the case of heterogeneous media, both ray theory (Hu *et al.* 2001) and spatially variable matched filters (Aoki & Schuster 2009) have been proposed to approximate PSFs. These were then used for migration deconvolution or 'deblurring' (Hu *et al.* 2001; Aoki & Schuster 2009), much similar to the pre-conditioning of a descent direction with the inverse approximate Hessian.

While the use of ray theory is restricted to well-defined and isolated high-frequency phases, the fitting of PSFs or their inverse with the help of matched filters could easily be generalized to full waveform inversion. In fact, our method proposed in Section 3 corresponds to a matched filter approach where one single input model is replaced by a set of input models that help to constrain the direction dependence of PSFs.

Our method for resolution analysis is, however, more general than resolution analysis is migration, mostly because it is applicable on all scales, to all types of waves and to any misfit functional that one finds suitable for a particular application.

7.4 The problem of multiple minima

The weak point of our analysis is the assumption that the global minimum of χ has been found. In large-scale applications this assumption can hardly be verified, and there is no sufficiently efficient optimization algorithm that does not risk being trapped in a local minimum. This restriction must be kept in mind when using the local Hessian in resolution analysis.

8 CONCLUSIONS

We have developed a new method for the quantitative resolution analysis in full seismic waveform inversion that overcomes the limitations of chequer board type synthetic inversions while being computationally more efficient. Our approach is based on the quadratic approximation of the misfit functional and the parametrization of the Hessian operator in terms of parent functions and their successive derivatives. The space-dependent parameters can be computed efficiently using Fourier transforms of the Hessian for a small set of linearly independent wavenumber vectors.

In the simplest case where the Hessian is parametrized in terms of Gaussians, we can infer 3-D distributions of direction-dependent resolution length and the distortion introduced by the tomographic method. Our approach can be considered a generalization of the ray density tensor (Kissling 1988) that quantifies the space-dependent azimuthal coverage, and therefore serves as a proxy for resolution in ray tomography. The advantages of the parametrized Hessian compared to the ray density tensor include the applicability to any type of seismic wave, and the rigorous quantification of resolution, for instance, in terms of the resolution length. Furthermore, the parametrized Hessian may be used for covariance and extremal bounds analysis, as proposed by Meju & Sakkas (2007) and Meju (2009).

A curious aspect of our approach is that information on the Hessian, and thus on resolution and trade-offs, can be obtained from linearized synthetic inversions where the initial model differs from the optimal model by a sinusoidal perturbation. This suggests that synthetics inversions are, at least in this sense, a useful tool, despite their well-known deficiencies (Lévêque *et al.* 1993). Yet, they should never be limited to single chequer board tests that bear little useful information.

1618 A. Fichtner and J. Trampert

As a corollary to our developments, we propose several improvements to full waveform inversion techniques—all of which require further investigations that are beyond the scope of this work. These include a new family of Newton-like methods, a pre-conditioner for optimization schemes of the conjugate-gradient type, an approach to adaptive parametrization independent from ray theory and a strategy for objective functional design that aims at maximizing resolution.

The most desirable improvement to our method would be more flexible but still rapidly converging parametrization of the Hessian. This would allow us, for instance, to study trade-offs between model parameters of different physical nature.

While the examples given in Section 4 are very specific, the method itself is very general. It may, in particular, be adapted to smaller scale exploration scenarios and to wave-equation-based tomography techniques that employ, for instance, georadar (e.g. Ernst *et al.* 2007; Meles *et al.* 2010) or microwave data (e.g. Chew & Lin 1995; Kosmas & Rappaport 2006).

ACKNOWLEDGMENTS

We would like to thank our colleagues Moritz Bernauer, Christian Böhm, Cédric Legendre, Helle Pedersen and Florian Rickers for inspiring comments and discussions. We are particularly grateful to Theo van Zessen for maintaining The STIG and GRIT, its little brother. Numerous computations were done on the Huygens IBM p6 supercomputer at SARA Amsterdam. Use of Huygens was sponsored by the National Computing Facilities Foundation (NCF) under the project SH-161-09 with financial support from the Netherlands Organisation for Scientific Research (NWO). Furthermore, we would like to thank Gerard Schuster for making us aware of the relation between the Hessian and PSFs. The comments and thought-provoking questions of Yann Capdeville and an anonymous reviewer helped us to improve the text. Finally, Andreas Fichtner gratefully acknowledges Deutsche Bahn for all the endless delays that provided ample time to finish this manuscript.

REFERENCES

- Aoki, N. & Schuster, G.T., 2009. Fast least-squares migration with a deblurring filter, *Geophysics*, 74, WCA83–WCA93.
- Backus, G.E. & Gilbert, F., 1967. Numerical application of a formalism for geophysical inverse problems, *Geophys. J. R. astr. Soc.*, 13, 247–276.
- 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.
- Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am.* geophys. Un., 81, F897.
- Bendinelli, O., Parmeggiani, G. & Zavatti, F., 1987. An analytical approximation of the Hubble space telescope monochromatic point spread function, J. Astrophys. Astron., 8, 343–350.
- Bernauer, M., Fichtner, A. & Igel, H., 2009. Inferring earth structure from combined measurements of rotational and translational ground motions, *Geophysics*, 74, WCD41–WCD47.
- Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global traveltime tomography, *J. geophys. Res.*, 103, 30 055–30 078.
- 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.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**, 1411–1436.
- Boschi, L., 2003. Measures of resolution in global body wave tomography, *Geophys. Res. Lett.*, **30**, doi:10.1029/2003GL018222.
- Boschi, L., Ekström, G. & Kustowski, B., 2004. Multiple resolution surface wave tomography: the Mediterranean basin, *Geophys. J. Int.*, 157, 293–304.
- Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.*, **185**, 845–870.
- 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.
- Brossier, R., Operto, S. & Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, 75, R37–R46.

- Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**, 1457–1473.
- Cara, M., Lévèque, J.J. & Maupin, V., 1984. Density-versus-depth models from multimode surface waves, *Geophys. Res. Lett.*, **11**, 633–636.
- Charlier, C.V.L., 1905. Über die Darstellung willkürlicher Funktionen, Arkiv för Matematik, Astronomi och Fysik, 20(2), 1–35.
- Chavent, G. & Plessix, R.-E., 1999. An optimal true-amplitude least-squares prestack depth-migration operator, *Geophysics*, 64, 508–515.
- Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances, *Pure Appl. Geophys.* 168, 1527–1552, doi:10.1007/s00024-010-0240-8.
- 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.
- Chew, W.C. & Lin, J.H., 1995. A frequency-hopping approach for microwave imaging of large inhomogeneous bodies, *IEEE Microw. Guid. Wave Lett.*, 5, 439–441.
- Cramér, H., 1957. *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ.
- Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic nonlinear waveform inversion: application to real data, *Geophysics*, 55, 527–538.
- Debayle, E. & Sambridge, M., 2004. Inversion of massive surface wave data sets: model construction and resolution assessment, *J. geophys. Res.*, 109, B02316, doi:10.1029/2003JB002652.
- Del Brio, E.B., Niguez, T.M. & Perote, J., 2009. Gram-Charlier densities: a multivariate approach, *Quant. Finance*, 9, 855–868.
- Devaney, A.J., 1984. Geophysical diffraction tomography, *IEEE Trans. Geos. Remote Sens.*, 22, 3–13.
- Devilee, R.J.R., Curtis, A. & Roy-Chowdhury, K., 1999. An efficient, probabilistic neural network approach to solving inverse problems: inverting surface wave velocities for Eurasian crustal thickness, *J. geophys. Res.*, 104, 28 841–28 857.
- Dondi, F., Betti, A., Blo, G. & Bighi, C., 1981. Statistical analysis of gaschromatographic peaks by the Gram-Charlier series of type A and the Edgeworth-Cramer series, *Anal. Chem.*, **53**, 496–504.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, 25, 297–356.
- 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.
- Ernst, J.R., Green, A.G., Maurer, H. & Holliger, K., 2007. Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data, *Geophysics*, **72**, J53–J64.

- Fang, Y., Cheney, M. & Roecker, S., 2010. Imaging from sparse measurements, *Geophys. J. Int.*, 180, 1289–1302.
- Ferreira, A.M.G. & Igel, H., 2009. Rotational motions of seismic surface waves in a laterally heterogeneous Earth, *Bull. seism. Soc. Am.*, 99, 1429–1436.
- 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. & Trampert, J., 2011. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.*, 185, 775– 798.
- 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 past states of the Australasian upper mantle, *Earth planet. Sci. Lett.*, 290, 270–280.
- Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics*, 51, 1387–1403.
- Gee, L.S. & Jordan, T.H., 1992. Generalized seismological data functionals, *Geophys. J. Int.*, 111, 363–390.
- Gram, J.P., 1883. Ueber die Entwickelung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate, *Journal f
 ür die reine und* angewandte Mathematik, 94, 6–73.
- Hingee, M., Tkalčić, H., Fichtner, A. & Sambridge, M., 2011. Seismic moment tensor inversion using a 3-D structural model: applications for the Australian region, *Geophys. J. Int.*, **184**, 949–964.
- Hu, J., Schuster, G.T. & Valasek, P.A., 2001. Poststack migration deconvolution, *Geophys. J. Int.*, 66, 939–952.
- 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.
- Jaynes, E.T., 2003. *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge.
- Kawai, K. & Geller, R.J., 2010. Waveform inversion for localised seismic structure and an application to D'' structure beneath the Pacific, J. geophys. Res., 115, doi:10.1029/2009JB006503.
- Kissling, E., 1988. Geotomography with local earthquake data, *Rev. Geo-phys.*, 26, 659–698.
- Konishi, K., Kawai, K., Geller, R.J. & Fuji, N., 2009. MORB in the lowermost mantle beneath the western Pacific: evidence from waveform inversion, *Earth planet. Sci. Lett.*, 278, 219–225.
- Kosmas, P. & Rappaport, C.M., 2006. FDTD-based time reversal for microwave breast cancer detection: localization in three dimensions, *IEEE Trans. Microw. Theo. Tech.*, **54**, 1921–1927.
- van Leeuwen, T. & Mulder, W.A., 2010. A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophys. J. Int.*, 182, 1383–1394.
- 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. seismol. Soc. Am., 96, 2383–2397.
- Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics*, 56, 645–653.
- Maurer, H., Greenhalgh, S. & Latzel, S., 2009. Frequency and spatial sam-

pling strategies for crosshole seismic waveform spectral inversion experiments, *Geophysics*, **74**, WCC11–WCC21.

- 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.
- Meles, G.A., van der Kruk, J., Greenhalgh, S.A., Ernst, J.R., Maurer, H. & Green, A.G., 2010. A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination of crosshole/borehole-to-surface GPR data, *IEEE Trans. Geosc. Rem. Sens.*, 48, 3391–3407.
- 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.
- Nemeth, T., Wu, C. & Schuster, G.T., 1999. Least-squares migration of incomplete reflection data, *Geophysics*, 64, 208–221.
- Nolet, G. & Montelli, R., 2005. Optimal parametrization of tomographic models, *Geophys. J. Int.*, 161, 365–372.
- Nolet, G., Montelli, R. & Virieux, J., 1999. Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, **138**, 36–44.
- Pankajakshan, P., Zhang, B., Blanc-Feraud, L., Kam, Z., Olivo-Marin, J.C. & Zerubia, J., 2000. Blind deconvolution for thin-layered confocal imaging, *Appl. Opt.*, 48, 4437–4448.
- Pavese, A., Artioli, G. & Prencipe, M., 1995. X-ray single-crystal diffraction study of pyrope in the temperature range 30–973 k, *Am. Mineral.*, 80, 457–464.
- Peter, D., Boschi, L., Deschamps, F., Fry, B., Ekström, G. & Giardini, D., 2008. A new finite-frequency shear-velocity model of the European-Mediterranean region, *Geophys. Res. Lett.*, 35, doi:10. 1029/2008GL034769.
- 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., Shin, C. & Hicks, G., 1998. Gauss-Newton and full Newton methods in frequency domain seismic waveform inversion, *Geophys. J. Int.*, 133, 341–362.
- Ravaut, C., Operto, S., Importa, L., Virieux, J., Herrero, A. & dell'Aversana, P., 2004. Multi-scale imaging of complex structures from multi-fold wide aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt, *Geophys. J. Int.*, **159**, 1032–1056.
- Ritsema, J. & van Heijst, H., 2002. Constraints on the correlation of *P* and *S*-wave velocity heterogeneity in the mantle from *P*, *PP*, *PPP* and *PKP* ab traveltimes, *Geophys. J. Int.*, **149**, 482–489.
- Ritsema, J., vanHeijst, H. & Woodhouse, J.H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, 286, 1925–1928.
- Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, 40, doi:10.1029/2000RG000089.
- Samuelson, P.A., 1943. Fitting general Gram-Charlier series, Ann. Math. Stat., 14(2), 179–187.
- Santosa, F. & Symes, W.W., 1988. Computation of the Hessian for leastsquares solutions of inverse problems of reflection seismology, *Inverse Probl.*, 4, 211–233.
- Sargan, J.D., 1975. Gram-Charlier approximations applied to t ratios of k-class estimators, *Econometrica*, 43, 327–347.

© 2011 The Authors, GJI, **187**, 1604–1624

- Sauer, P.W. & Heydt, G.T., 1979. Convenient multivariate Gram-Charlier type-A series, *IEEE Trans. Comm.*, 27, 247–248.
- Schäfer, J., Boschi, L. & Kissling, E., 2011. Adaptively parametrized surface wave tomography: methodology and a new model of the European upper mantle, *Geophys. J. Int.*, **186**, 1431–1453.
- Schivardi, R. & Morelli, A., 2011. EPmantle: a 3-D transversely isotropic model of the upper mantle under the European Plate, *Geophys. J. Int.*, 185, 469–484.
- Schuster, G.T. & Hu, J., 2000. Green's function for migration: continuous recording geometry, *Geophysics*, 65, 167–175.
- Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics*, 69, 231– 248.
- Smithyman, B., Pratt, R.G., Hayles, J. & Wittebolle, R., 2009. Detecting near-surface objects with seismic waveform tomography, *Geophysics*, 74, WCC119–WCC127.
- Spakman, W., 1991. Delay-time tomography of the upper mantle below Europe, the Mediterranean and Asia Minor, *Geophys. J. Int.*, 107, 309– 332.
- Takam Takougang, E.M. & Calvert, A., 2011. Application of waveform tomography to marine seismic reflection data from the Queen Charlotte Basin of western Canada, *Geophysics*, 76, B55–B70.
- Takeuchi, H. & Saito, M., 1972. Seismic surface waves, in *Methods in Computational Physics*, Vol. 11, pp. 217–295, ed. Bolt B.A., Academic Press, New York, NY.
- 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., 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.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, 160, 195–216.
- Trujillo, I., Aguerri, J.A.L. & Gutierrez, C.M., 2001. The effects of seeing on Sérsic profiles, *Mon. Not. R. astr. Soc.*, **321**, 269–276.
- Widiyantoro, S. & van der Hilst, R., 1997. Mantle structure beneath Indonesia inferred from high-resolution tomographic imaging, *Geophys. J. Int.*, 130, 167–182.
- Wu, R.-S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics*, 52, 11–25.
- Yu, J.H., Wang, Y. & Schuster, G.T., 2002. Deconvolution of the psf of a seismic lens, in *Image Reconstruction from Incomplete Data II*, Proc. SPIE Vol. 4792, pp. 135–145, eds Bones, P.J., Fiddy, M.A. & Millane, R.P., SPIE, Bellingham, WA.
- Zhang, B., Zerubia, J. & Olivo-Marin, J.C., 2007. Gaussian approximation of fluorescence microscope point-spread function models, *Appl. Opt.*, 46, 1819–1829.
- Zielhuis, A. & Nolet, G., 1994. Shear-wave velocity variations in the upper mantle beneath central Europe, *Geophys. J. Int.*, **117**, 695–715.

APPENDIX A: HESSIAN KERNELS

In this appendix, we provide a condensed recipe for the efficient computation of the Hessian applied to arbitrary model perturbations. In the interest of notational compactness, we introduce the notion of Hessian kernels $h(\delta m; x)$, defined as the Hessian H(x, y) applied to a model perturbation $\delta m(x)$.

$$\mathbf{h}(\delta \mathbf{m}; \mathbf{x}) = \int_{G} \mathbf{H}(\mathbf{x}, \mathbf{y}) \, \delta \mathbf{m}(\mathbf{y}) \, \mathrm{d}^{3} \mathbf{y} \,. \tag{A1}$$

More details including numerous examples may be found in Fichtner & Trampert (2011).

A1 The computation of Hessian kernels using adjoint techniques

As a preparatory step, we express the Fréchet derivative of the misfit functional, $\nabla_m \chi \, \delta \mathbf{m}$, in terms of the Fréchet kernel $\mathbf{K}(\mathbf{x})$

$$\nabla_m \chi \,\delta \mathbf{m} = \int_G \delta \mathbf{m}(\mathbf{x})^T \,\mathbf{K}(\mathbf{x}) \,\mathrm{d}^3 \mathbf{x} \,. \tag{A2}$$

In the exemplary case of an isotropic medium parametrized in terms of the *S* velocity β , the *P* velocity α and density ρ , the Fréchet kernel has three components, $\mathbf{K} = (K_{\beta}, K_{\alpha}, K_{\rho})^{T}$, and eq. (A2) can be expanded to

$$\nabla_m \chi \,\delta \mathbf{m} = \int_G K_\beta(\mathbf{x}) \,\delta\beta(\mathbf{x}) \,\mathrm{d}^3 \mathbf{x} + \int_G K_\alpha(\mathbf{x}) \,\delta\alpha(\mathbf{x}) \,\mathrm{d}^3 \mathbf{x} + \int_G K_\rho(\mathbf{x}) \,\delta\rho(\mathbf{x}) \,\mathrm{d}^3 \mathbf{x} \,. \tag{A3}$$

The Fréchet kernel can be computed efficiently with the help of adjoint techniques (e.g. Tarantola 1988; Tromp *et al.* 2005; Fichtner *et al.* 2006; Liu & Tromp 2006; Plessix 2006; Chen 2011) that express **K** in terms of the current earth model $\mathbf{m}(\mathbf{x})$, the forward wave field $\mathbf{u}(\mathbf{x}, t)$ and the adjoint wave field $\mathbf{u}^{\dagger}(\mathbf{x}, t)$, that is,

$$\mathbf{K}(\mathbf{x}) = \mathbf{K}(\mathbf{m}, \mathbf{u}, \mathbf{u}^{\mathsf{T}}; \mathbf{x}). \tag{A4}$$

The semicolon in eq. (A4) indicates that the variables to its left are fixed for a given realization of the space-dependent kernel **K**. We will adhere to this semicolon convention throughout the following developments. The adjoint wave field $\mathbf{u}^{\dagger}(\mathbf{x}, t)$ is the solution of the adjoint wave equation, symbolically written in terms of the adjoint \mathbf{L}^{\dagger} of the wave equation operator **L**

$$\mathbf{L}^{\dagger}(\mathbf{m},\mathbf{u}^{\dagger}) = \mathbf{f}^{\dagger}(\mathbf{m}). \tag{A5}$$

The adjoint source \mathbf{f}^{\dagger} acts at the receiver positions, and its time evolution is completely determined by the misfit χ . This explains the dependence of \mathbf{f}^{\dagger} on the earth model \mathbf{m} . Note that both the forward and the adjoint wave fields, \mathbf{u} and \mathbf{u}^{\dagger} , are also dependent on \mathbf{m} , that is, $\mathbf{u} = \mathbf{u}(\mathbf{m}; \mathbf{x})$ and $\mathbf{u}^{\dagger} = \mathbf{u}^{\dagger}(\mathbf{m}; \mathbf{x})$

To obtain a relation between the Fréchet kernel \mathbf{K} , the Hessian kernel \mathbf{h} and the Hessian operator \mathbf{H} , we differentiate eq. (A2) once more with respect to \mathbf{m} :

$$\nabla_m \nabla_m \chi \,\delta \mathbf{m} \,\delta \mathbf{m} = \int_G \delta \mathbf{m}(\mathbf{x})^T \left[\nabla_m \mathbf{K}(\mathbf{x}) \,\delta \mathbf{m} \right] \,\mathrm{d}^3 \mathbf{x} = \int_G \int_G \delta \mathbf{m}(\mathbf{x})^T \,\mathbf{H}(\mathbf{x}, \mathbf{y}) \,\delta \mathbf{m}(\mathbf{y}) \,\mathrm{d}^3 \mathbf{x} \,\mathrm{d}^3 \mathbf{y} \,. \tag{A6}$$

It follows from (A6) that the Hessian kernel **h** from eq. (A2) is equal to the total Fréchet derivative of the Fréchet kernel **K** in the direction of $\delta \mathbf{m}$

$$\mathbf{h}(\delta \mathbf{m}; \mathbf{x}) = \nabla_m \mathbf{K}(\mathbf{x}) \,\delta \mathbf{m} \,. \tag{A7}$$

To compute **h** explicitly, we exploit the dependence of the Fréchet kernel on **m**, **u** and \mathbf{u}^{\dagger} . The application of the chain rule to eq. (A7) then yields

$$\mathbf{h}(\delta \mathbf{m}; \mathbf{x}) = \nabla_m \mathbf{K}(\mathbf{x}) \,\delta \mathbf{m} = \partial_m \mathbf{K}(\mathbf{m}, \mathbf{u}, \mathbf{u}^{\dagger}; \mathbf{x}) \,\delta \mathbf{m} + \mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger}; \mathbf{x}) + \mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger}; \mathbf{x}) \,, \tag{A8}$$

where $\partial_m \mathbf{K} \, \delta \mathbf{m}$ denotes the partial Fréchet derivative of \mathbf{K} in the direction $\delta \mathbf{m}$; as opposed to the total Fréchet derivative $\nabla_m \mathbf{K} \, \delta \mathbf{m}$. The symbols $\delta_m \mathbf{u} = \nabla_m \mathbf{u} \, \delta \mathbf{m}$ and $\delta_m \mathbf{u}^{\dagger} = \nabla_m \mathbf{u}^{\dagger} \, \delta \mathbf{m}$ are the Fréchet derivatives of the forward and adjoint fields, respectively.

Eq. (A8) reveals that the Hessian kernel $\mathbf{h}(\delta \mathbf{m}; \mathbf{x})$ has three contributions. The first one of these, $\partial_m \mathbf{K}(\mathbf{m}, \mathbf{u}, \mathbf{u}^{\dagger}; \mathbf{x}) \delta \mathbf{m}$, is localized at the perturbation $\delta \mathbf{m}$, and has no obvious physical interpretation. The second contribution, $\mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger}; \mathbf{x})$, closely resembles the Fréchet kernel from eq. (A4). However, instead of \mathbf{u} it involves the derivative $\delta_m \mathbf{u}$, that can be interpreted as the scattered forward field that is excited when \mathbf{u} impinges upon the perturbation $\delta \mathbf{m}$. In practice, $\delta_m \mathbf{u}$ can be computed most conveniently using finite-differences such as

$$\delta_m \mathbf{u} \approx \frac{1}{\nu} [\mathbf{u}(\mathbf{m} + \nu \, \delta \mathbf{m}) - \mathbf{u}(\mathbf{m})], \tag{A9}$$

where ν must be sufficiently small to ensure the validity of the first-order approximation. From a physical point of view, this second contribution to the Hessian kernel describes the effect of second-order scattering from an arbitrary perturbation $\delta \mathbf{m}'$ to $\delta \mathbf{m}$ and from there to the receiver.

The third contribution to the Hessian kernel, $\mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger}; \mathbf{x})$, also resembles the Fréchet kernel (A4), but instead of the adjoint field \mathbf{u}^{\dagger} , it involves the derivative $\delta_m \mathbf{u}^{\dagger}$. While $\delta_m \mathbf{u}^{\dagger}$ may again be approximated with finite differences, for example,

$$\delta_m \mathbf{u}^{\dagger} \approx \frac{1}{\nu} [\mathbf{u}^{\dagger} (\mathbf{m} + \nu \, \delta \mathbf{m}) - \mathbf{u}^{\dagger} (\mathbf{m})], \tag{A10}$$

its computation and interpretation are more complex than it may seem. In fact, a model perturbation from **m** to $\mathbf{m} + \delta \mathbf{m}$ has two effects: (1) the adjoint field \mathbf{u}^{\dagger} is scattered when it impinges upon $\delta \mathbf{m}$ and (2) the adjoint source \mathbf{f}^{\dagger} changes as well because a model perturbation affects the misfit χ that determines \mathbf{f}^{\dagger} . It follows that the computation of $\mathbf{u}^{\dagger}(\mathbf{m} + \nu \, \delta \mathbf{m})$ requires the solution of the adjoint wave equation for the perturbed earth model and the perturbed adjoint source, that is,

$$\mathbf{L}^{\dagger}(\mathbf{m} + \nu \,\delta \mathbf{m}, \mathbf{u}^{\dagger}) = \mathbf{f}^{\dagger}(\mathbf{m} + \nu \,\delta \mathbf{m}) \,. \tag{A11}$$

In contrast to the second contribution $\mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger}; \mathbf{x})$, the third contribution $\mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger}; \mathbf{x})$ accounts for both the first-order scattering that leads to the perturbation of the adjoint source and the second-order scattering between $\delta \mathbf{m}$ and the receiver.

A2 A realistic example

To attach physical intuition to the above equations, we consider an isotropic medium parametrized in terms of the *P* velocity α , the *S* velocity β and density ρ , that is, $\mathbf{m} = (\alpha, \beta, \rho)^T$. The Fréchet kernel K_β is given by (e.g. Tromp *et al.* 2005; Liu & Tromp 2006; Fichtner 2010)

$$K_{\beta} = 4\rho\beta \int \epsilon_{ij} \epsilon_{ij}^{\dagger} dt - 4\rho\alpha \int \epsilon_{ii} \epsilon_{jj}^{\dagger} dt , \qquad (A12)$$

where $\epsilon^{\dagger} = \frac{1}{2} [\nabla \mathbf{u}^{\dagger} + (\nabla \mathbf{u}^{\dagger})^{T}]$ is the strain tensor computed from the adjoint field. For the specific case of a 25 s Love wave recorded at 25.2° epicentral distance, the Fréchet kernel for a cross-correlation time shift measurement (Luo & Schuster 1991) occupies a cigar-shaped volume that extends from the source to the receiver (Fig. A1).

Making use of eq. (A8) we find that the β -component of the Hessian kernel corresponding to an *S* velocity perturbation $\delta \mathbf{m} = (0, \delta \beta, 0)^T$, that is, $h_\beta(\delta \beta) = \int H_{\beta\beta} \,\delta \beta \,d^3 \mathbf{y}$, is given by

$$h_{\beta}(\delta\beta) = \underbrace{4\rho \,\delta\beta \int \epsilon_{ij} \,\epsilon_{ij}^{\dagger} \,dt}_{\partial_m \mathbf{K}(\mathbf{m}, \mathbf{u}, \mathbf{u}^{\dagger}) \,\delta\mathbf{m}} + \underbrace{4\rho\beta \int \delta_{\beta} \epsilon_{ij} \,dt - 4\rho\alpha \int \delta_{\beta} \epsilon_{ii} \,\epsilon_{jj}^{\dagger} \,dt}_{\mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger})} + \underbrace{4\rho\beta \int \epsilon_{ij} \,\delta_{\beta} \epsilon_{ij}^{\dagger} \,dt - 4\rho\alpha \int \epsilon_{ii} \,\delta_{\beta} \epsilon_{jj}^{\dagger} \,dt}_{\mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger})} .$$
(A13)

The contribution $\partial_m \mathbf{K}(\mathbf{m}, \mathbf{u}, \mathbf{u}^{\dagger}) \, \delta \mathbf{m}$ can be computed directly from the Fréchet kernels. Additional simulations are needed for $\mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger})$ and $\mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger})$ because these involve the differentiated forward and adjoint fields, $\delta \mathbf{u}$ and $\delta \mathbf{u}^{\dagger}$, that can be computed most conveniently with the help of the finite-difference approximations (A9) and (A10).

The three contributions to the Hessian kernel $h_{\beta}(\delta\beta)$ for a point perturbation $\delta\beta$ located beneath southern Sardinia are shown separately in the top row of Fig. A2. The contribution $\mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger})$ is a superposition of two components, labelled *F* and *S*. These correspond to the primary influence zone represented by the approximate Hessian (*F*) and to the secondary influence zone where second-order scattering affects the measurement. The contribution $\mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger})$ only extends from the receiver to $\delta\beta$. Second-order scattering from $\delta\beta$ to an *S* velocity



Figure A1. (Left) 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) Fréchet kernel \bar{K}_{vs} for a cross-correlation time shift measurement on the N–S-component Love wave shown to the left. The source and receiver positions are marked by + and •, respectively. Figure modified from Fichtner & Trampert (2011).



Figure A2. Hessian kernels plotted at the surface of the Earth. The position of the *S* velocity perturbation $\delta\beta$ is indicated by \circ . Top row: The individual contributions $\mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger})$, $\partial_m \mathbf{K}(\mathbf{m}, \mathbf{u}, \mathbf{u}^{\dagger}) \delta \mathbf{m}$ and $\mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger})$ (from left to right). The contribution $\mathbf{K}(\mathbf{m}, \mathbf{u}, \delta_m \mathbf{u}^{\dagger})$ has two components, labelled *F* and *S*. These represent a primary influence zone corresponding to the approximate Hessian (*F*), and to a secondary influence zone where second-order scattering affects the measurement (*S*). Bottom row: The complete Hessian kernel $h_{\beta}(\delta\beta)$. The contribution from second-order scattering is again encircled by a dotted curve and labelled *S*. Figure modified from Fichtner & Trampert (2011).

perturbation within the non-zero parts of $\mathbf{K}(\mathbf{m}, \delta_m \mathbf{u}, \mathbf{u}^{\dagger})$ has an effect on the measurement. Finally, $\partial_m \mathbf{K}(\mathbf{m}, \mathbf{u}, \mathbf{u}^{\dagger}) \delta \mathbf{m}$ is restricted to the volume occupied by $\delta \beta$. The complete Hessian kernel $h_{\beta}(\delta \beta)$ is shown in the bottom row of Fig. A2. It can be interpreted as the continuous representation of the row of the Hessian matrix that corresponds to the basis function coincident with $\delta \beta$. The fully discrete row is obtained by projecting $h_{\beta}(\delta \beta)$ onto the basis functions. For a more detailed discussion of Hessian kernels, the reader is referred to Fichtner & Trampert (2011).

APPENDIX B: MULTIVARIATE GRAM-CHARLIER SERIES

The resolution analysis proposed in Section 3 rests on the approximation of the Hessian H(x, y) by a linear sum of terms that consist of a parent function and its successive derivatives. Approximation problems of this type have been studied at least since the end of the 19th century (e.g. Gram 1883; Charlier 1905; Samuelson 1943), and the special case where the parent function is a Gaussian became known the Gram-Charlier series. While originally introduced in a purely mathematical context, Gram-Charlier series became a popular tool for the approximation of 1-D densities in a variety of fields including econometrics (e.g. Sargan 1975), analytical chemistry (e.g. Dondi *et al.* 1981) and mineral physics (e.g. Pavese *et al.* 1995). Generalizations to higher dimensions and applications in a geophysical context have, however, remained an exception (e.g. Sauer & Heydt 1979; Gee & Jordan 1992; Del Brio *et al.* 2009). In the following paragraphs, we describe Gram-Charlier approximations in $N \ge 1$ dimensions that are based on the representation of a function in terms of its cumulants.

B1 Moments and cumulants

We consider the spatial Fourier transform of a function $f(\mathbf{x})$, defined as

$$\widetilde{f}(\mathbf{k}) = \int_{\mathbb{R}^N} f(\mathbf{x}) e^{i\mathbf{k}^T \mathbf{x}} d^N \mathbf{x}.$$

(B1)

Expanding $\tilde{f}(\mathbf{x})$ into a Taylor series, yields

$$\widetilde{f}(\mathbf{k}) = M^{(0)} + i\mathbf{k}^T \mathbf{M}^{(1)} + \frac{i^2}{2!} \mathbf{k}^T \mathbf{M}^{(2)} \mathbf{k} + \cdots,$$
(B2)

where the components of the tensorial Taylor coefficients $\mathbf{M}^{(n)}$ are given by

$$M_{ijk\dots}^{(n)} = \frac{1}{i^n} \left(\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_k} \dots \right) \widetilde{f}|_{\mathbf{k}=\mathbf{0}} .$$
(B3)

Introducing (B1) into (B3), reveals that the coefficients $M_{ijk...}^{(n)}$ are equal to the moments of f.

$$M_{ijk\dots}^{(n)} = \int_{\mathbb{R}^N} (x_i x_j x_k \dots) f(\mathbf{x}) d^N \mathbf{x}.$$
(B4)

Similar to (B2), the cumulants $\mathbf{F}^{(n)}$ of f are defined through the Taylor coefficients of log \tilde{f} .

$$\log \widetilde{f}(\mathbf{k}) = F^{(0)} + i\mathbf{k}^T \mathbf{F}^{(1)} + \frac{i^2}{2!} \mathbf{k}^T \mathbf{F}^{(2)} \mathbf{k} + \cdots$$
(B5)

Upon using

$$F_{ijk\dots}^{(n)} = \frac{1}{i^n} \left(\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_k} \dots \right) \log \tilde{f}|_{\mathbf{k}=\mathbf{0}},$$
(B6)

we find that the first three cumulants and moments are related as follows:

$$F^{(0)} = \log M^{(0)}, \quad F_i^{(1)} = \frac{M_i^{(1)}}{M^{(0)}}, \quad F_{ij}^{(2)} = \frac{1}{M^{(0)}} \left(M_{ij}^{(2)} - \frac{M_i^{(1)} M_j^{(1)}}{M^{(0)}} \right).$$
(B7)

By definition (eq. B5), the cumulants are additive under convolution.

B2 Wavenumber-domain Gram-Charlier series

Our goal is to approximate f in terms of a suitably chosen parent function g. With the help of eq. (B5), we can express the spectrum \tilde{f} as

$$\widetilde{f}(\mathbf{k}) = \frac{f(\mathbf{k})}{\widetilde{g}(\mathbf{k})} \widetilde{g}(\mathbf{k}) = \exp\left[\left(F^{(0)} - G^{(0)}\right) + i\mathbf{k}^T \left(\mathbf{F}^{(1)} - \mathbf{G}^{(1)}\right) + \frac{i^2}{2!} \mathbf{k}^T \left(\mathbf{F}^{(2)} - \mathbf{G}^{(2)}\right) \mathbf{k} + \cdots\right] \widetilde{g}(\mathbf{k}),$$
(B8)

where $\mathbf{F}^{(n)}$ and $\mathbf{G}^{(n)}$ are the cumulants of f and g, respectively. Eq. (B8) provides a very general expression of \tilde{f} , that can be greatly simplified through the judicious choice of g. Thus, in the context of resolution analysis, the parent function should already capture the principal characteristics of the PSF. We obtain the wavenumber-domain representation of the classical Gram–Charlier A series when the parent function g is equal to the normalized Gaussian

$$g(\mathbf{x}) = \frac{\sqrt{\det \mathbf{C}}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2} \left(\mathbf{x} - \mathbf{z}\right)^T \mathbf{C} \left(\mathbf{x} - \mathbf{z}\right)\right]$$
(B9)

with the cumulants

$$G^{(0)} = 0$$
, $\mathbf{G}^{(1)} = \mathbf{z}$, $\mathbf{G}^{(2)} = \mathbf{C}^{-1}$, $\mathbf{G}^{(n>2)} = \mathbf{0}$. (B10)

This eliminates the cumulants of g with orders greater than 2 from eq. (B8). Since we are still free to choose the specific values of $\mathbf{G}^{(1)}$ and $\mathbf{G}^{(2)}$, we let $\mathbf{G}^{(1)} = \mathbf{F}^{(1)}$ and $\mathbf{G}^{(2)} = \mathbf{F}^{(2)}$, so that eq. (B8) reduces to

$$\widetilde{f}(\mathbf{k}) = M^{(0)} \exp\left[\frac{i^3}{3!} F^{(3)}_{ijk} k_i k_j k_k + \frac{i^4}{4!} F^{(4)}_{ijkl} k_i k_j k_k k_l + \cdots\right] \exp\left(i\mathbf{k}^T \mathbf{z} - \frac{1}{2} \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k}\right),$$
(B11)

where $M^{(0)}$ is the zeroth moment of f. Eq. (B11) expresses the spectrum of f in terms of its cumulants with orders $n \ge 3$, and its first three moments: the norm $M^{(0)}$, the expected value z and the covariance C. By expanding the first exponential on the left-hand side of (B11) into a Taylor series, we obtain the following approximation:

$$\widetilde{f}(\mathbf{k}) = M^{(0)} \left[1 + \frac{i^3}{3!} F^{(3)}_{ijk} k_i k_j k_k + \frac{i^4}{4!} F^{(4)}_{ijkl} k_i k_j k_k k_l + \cdots \right] \exp\left(-i\mathbf{k}^T \mathbf{z} - \frac{1}{2} \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k}\right).$$
(B12)

While the coefficients of orders 3, 4 and 5 are equal to $\mathbf{F}^{(3)}$, $\mathbf{F}^{(4)}$ and $\mathbf{F}^{(5)}$, respectively, the higher order coefficients involve products of cumulants.

B3 Space-domain Gram-Charlier series

To translate the wavenumber-domain Gram-Charlier series (B11) into the space domain, we exploit the Fourier correspondence

$$\exp\left(ik_{j}\right) \widetilde{f}(\mathbf{k}) \leftrightarrow \exp\left(-\frac{\partial}{\partial x_{j}}\right) f(\mathbf{x}). \tag{B13}$$



Figure B1. (Left) Real part of the step function spectrum $\tilde{f}(k)$ from eq. (B17), plotted in black. The Gram–Charlier approximations of orders 0, 3 and 7 are shown in different red tones. (Centre) The same as to the left, but for the imaginary part of $\tilde{f}(k)$. (Right) The space-domain step function from (B17) and its Gram–Charlier approximations of orders 0, 3 and 7.

With the help of (B13), eq. (B11) transforms to

$$f(\mathbf{x}) = M^{(0)} \frac{\sqrt{\det \mathbf{C}}}{(2\pi)^{N/2}} \exp\left[\frac{(-1)^3}{3!} F_{ijk}^{(3)} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_k} + \frac{(-1)^4}{4!} F_{ijkl}^{(4)} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_k} \frac{\partial}{\partial x_l} + \cdots \right] \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{z})^T \mathbf{C} (\mathbf{x} - \mathbf{z})\right].$$
(B14)

Again expanding the first exponential on the left-hand side of (B14), yields a more useful equation

$$f(\mathbf{x}) = M^{(0)} \frac{\sqrt{\det \mathbf{C}}}{(2\pi)^{N/2}} \left[1 + \frac{(-1)^3}{3!} F_{ijk}^{(3)} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_k} + \frac{(-1)^4}{4!} F_{ijkl}^{(4)} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_k} \frac{\partial}{\partial x_l} + \cdots \right] \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{z})^T \mathbf{C} (\mathbf{x} - \mathbf{z}) \right].$$
(B15)

The Gram-Charlier series from (B15) can be interpreted as the summation of contributions with increasing wavelength. The truncated expansion

$$f(\mathbf{x}) \approx M^{(0)} \frac{\sqrt{\det \mathbf{C}}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2} \left(\mathbf{x} - \mathbf{z}\right)^T \mathbf{C} \left(\mathbf{x} - \mathbf{z}\right)\right]$$
(B16)

is, in this sense, a long-wavelength approximation.

B4 Properties of the Gram-Charlier expansion

To illustrate the nature of Gram-Charlier expansion, we consider the step function

$$f(x) = \begin{cases} 1, & -2 \le x < 0, \\ \frac{1}{2}, & 0 \le x \le 2, \\ 0, & \text{otherwise}, \end{cases}$$

$$\widetilde{f}(k) = [2 \exp(-ik) + \exp(ik)] \operatorname{sinc}(k)$$
(B17)

as a simple 1-D example. The wavenumber- and space-domain Gram–Charlier expansions of orders 0, 3 and 7 are shown in Fig. B1. The 0th-order approximation in the wavenumber domain is a Gaussian that closely matches the exact spectrum $\tilde{f}(k)$ within the wavenumber range $\pm 1 \text{ m}^{-1}$. The corresponding space-domain approximation is a Gaussian where the first three moments are equal to those of the step function (B17). Higher order contributions account for the skewness of f(x), but the improvements are limited to the wavenumber interval $[-\pi, \pi]$, outside of which the Taylor series from eq. (B11) does not converge. It follows that the Gram–Charlier expansion of (B17) is always a low-wavenumber approximation, even when infinitely many coefficients are included. This behaviour is typical in the sense that a Gram–Charlier series does not necessarily converge, and if it does, the convergence may be slow (e.g. Samuelson 1943; Cramér 1957).

The essence of this example is that the Gram–Charlier series should be considered a tool to capture the gross features of a function with a few coefficients, where the meaning of gross feature is determined by the choice of the parent function.