The impact of approximations and arbitrary choices on geophysical images

Andrew P. Valentine and Jeannot Trampert

Department of Earth Sciences, Universiteit Utrecht, P.O. Box 80.115, 3508*TC Utrecht, The Netherlands. E-mail: a.p.valentine@uu.nl*

SUMMARY

Whenever a geophysical image is to be constructed, a variety of choices must be made. Some, such as those governing data selection and processing, or model parametrization, are somewhat arbitrary: there may be little reason to prefer one choice over another. Others, such as defining the theoretical framework within which the data are to be explained, may be more straightforward: typically, an 'exact' theory exists, but various approximations may need to be adopted in order to make the imaging problem computationally tractable. Differences between any two images of the same system can be explained in terms of differences between these choices. Understanding the impact of each particular decision is essential if images are to be interpreted properly—but little progress has been made towards a quantitative treatment of this effect. In this paper, we consider a general linearized inverse problem, applicable to a wide range of imaging situations. We write down an expression for the difference between two images produced using similar inversion strategies, but where different choices have been made. This provides a framework within which inversion algorithms may be analysed, and allows us to consider how image effects may arise. In this paper, we take a general view, and do not specialize our discussion to any specific imaging problem or setup (beyond the restrictions implied by the use of linearized inversion techniques). In particular, we look at the concept of 'hybrid inversion', in which highly accurate synthetic data (typically the result of an expensive numerical simulation) is combined with an inverse operator constructed based on theoretical approximations. It is generally supposed that this offers the benefits of using the more complete theory, without the full computational costs. We argue that the inverse operator is as important as the forward calculation in determining the accuracy of results. We illustrate this using a simple example, based on imaging the density structure of a vibrating string.

Key words: Inverse theory; Numerical approximations and analysis; Tomography.

1 INTRODUCTION

Much research in the geosciences builds, to some degree, on information gleaned from geophysical images—models for earth systems that have been constructed based on observed data. Invariably, no two images of the same system agree (see e.g. the comparisons presented in Schaeffer & Lebedev 2013, 2015): some of this discrepancy may be attributable to the use of different data sets, but much is due to differences in the inversion framework used (Boschi & Dziewoński 1999). Each image is valid, in so far as each represents the model that best explains a given data set, within some particular theoretical basis, and subject to some set of assumptions. However, the extent to which these choices introduce biases or deficiencies into models is difficult to assess, presenting a barrier to robust interpretation.

Equally, this issue can be posed from the perspective of the researcher setting out to perform imaging: how does one achieve 'best' results? Of particular interest at present is the question of how best to use computationally expensive exact numerical approaches to modeling geophysical phenomena: is this always worth the cost, or could the computational resources be more effectively deployed elsewhere?

Most geophysical imaging approaches fall into one of two broad categories. The first, which will be the focus of this paper, encompasses algorithms built on linearized inverse theory: models are iteratively refined to bring synthetic (predicted) data into agreement with observations and the 'image' is a single model. The second category constitutes 'sampling-based approaches': a large collection of random models are tested against observations and the 'image' is typically a probability distribution for model parameters. This has the benefit that a fairly complete understanding of the relationship between models and data can be obtained; however, large-scale applications can be challenging due to the computational cost of testing each model. We will not discuss sampling approaches further;

examples include Sambridge (1999), Bodin & Sambridge (2009), Dosso *et al.* (2012), Käufl *et al.* (2014) and de Wit *et al.* (2014).

Within the framework of linear inverse theory, various approaches have been developed to allow quantification of the accuracy or quality of any particular inversion. Measures of resolution quantify the degree to which a hypothetical structural feature could be imaged using a particular technique: either through direct inversion of synthetic data (such as the 'chequerboard test', which remains popular despite its limitations—see Lévêque *et al.* 1993), or through analysis of the resolution operator (e.g. Trampert *et al.* 2013). The degree to which the imaging process constrains any particular model parameter may be assessed by exploring the relationship between model and synthetic in the region of the best-fitting model (e.g. Meju & Hutton 1992; Valentine & Trampert 2012), by relying on the assumptions underpinning linear inverse theory (e.g. Fichtner & Trampert 2011b), or by framing the inversion in a Bayesian context (e.g. Tarantola & Valette 1982). In principle, these techniques allow uncertainties or systematic errors associated with data noise and theoretical approximations to be taken into account. However, the results are limited by the validity of the imaging approach used, and the role of any assumptions underpinning this remain difficult to test.

In this paper, we develop and analyse a general expression describing the relationship between images produced by any two distinct linearized inversion algorithms. We begin by setting out a general formulation of linear inverse theory, and use this to obtain an expression for the difference between models produced by any pair of inversions. In particular, we consider the case where weakly non-linear problems are tackled iteratively using a linear algorithm. By analysing this expression, we can determine the various effects that contribute to discrepancies between geophysical images, and we can explore how certain common choices may affect images. In particular, we consider the case of 'hybrid' inversion schemes, whereby high-quality numerical forward calculations are combined with approximate inverse operators, and demonstrate that this approach does not necessarily yield any significant improvement over computing all quantities approximately. Our analysis may similarly apply to inversions based on the adjoint technique, particularly where smoothing or other post-processing is carried out. We also briefly consider a common modification to the inversion algorithm that takes the limited resolution of the imaging operator into account, and results in a non-linear inversion operator. We illustrate this discussion using a toy problem—imaging the density structure of a non-uniform vibrating string—which has a number of features in common with geophysical inversions for earth structure.

2 LINEARIZED INVERSE THEORY

Solving an inverse problem requires us to find the model, $\mathbf{m}$, for which synthetic data, $\mathbf{s}(\mathbf{m})$, are in closest agreement with observations, $\mathbf{d}$. For any points in model space, $\mathbf{m_i}$ and $\mathbf{m_{i+1}}$ and linear operator $\mathcal{S}_{\mathbf{m_i}}$, we can write the function $\mathbf{s}(\mathbf{m})$ in the form of a linear term and a remainder

$$\mathbf{s}(\mathbf{m_{i+1}}) = \mathbf{s}(\mathbf{m_i}) + \mathcal{S}_{\mathbf{m_i}}(\mathbf{m_{i+1}} - \mathbf{m_i}) + \mathbf{b}_\mathcal{S}(\mathbf{m_i}, \mathbf{m_{i+1}}) \quad (1)$$

where $\mathbf{b}_\mathcal{S}$ is defined by eq. (1), and encapsulates all functional behaviour not described by $\mathcal{S}_{\mathbf{m_i}}$. If we can identify a region of model space, $\mathbb{S}_{\mathbf{m_i}}$, such that

$$\|\mathbf{b}_\mathcal{S}(\mathbf{m_i}, \mathbf{m_{i+1}})\| < \varepsilon \; \forall \; \mathbf{m_{i+1}} \in \mathbb{S}_{\mathbf{m_i}} \quad (2)$$

for appropriately chosen ε, we regard $\mathcal{S}_{\mathbf{m_i}}$ as a good linear approximation to the behaviour of $\mathbf{s}(\mathbf{m})$ within that region. Motivated by

this, and identifying the predictions at $\mathbf{s}(\mathbf{m_{i+1}})$ with the data we seek to explain, we can express a general linear inversion algorithm in the simple form

$$\mathbf{m_{i+1}} = \mathbf{m_i} + \mathcal{S}_{\mathbf{m_i}}^{-g}(\mathbf{d} - \mathbf{s}(\mathbf{m_i})) \quad (3)$$

where $\mathcal{S}^{-g}$ represents some general inverse operator; we assume that it, too, is a linear operator. The choice of notation here suggests that $\mathcal{S}^{-g}$ is somehow constructed from $\mathcal{S}$. In most practical cases this will be true, and the inverse is usually derived by minimization of some specified 'misfit function', which quantifies the agreement between observations and synthetic data, under the assumption that eq. (1) is valid, and the $\mathbf{b}_\mathcal{S}$ term negligible. However, the discussion in this paper does not require us to place any restrictions on the relationship between forward and inverse operators. It will typically be beneficial to apply eq. (3) iteratively, updating predictions and linear operators to reflect the improved model.

Fundamentally, this paper centres on understanding how changes to $\mathcal{S}^{-g}$ and $\mathbf{s}$ are reflected in the image recovered at any given iteration. In general, we are concerned with deliberate changes, associated with choices made in establishing the inversion framework. However, we draw readers' attention to the existence of a large body of work in the numerical analysis literature, focusing on the stability of linear systems (e.g. van der Sluis 1975; Stewart 1977; Golub & van Loan 2012). In these papers, interest is typically focused on 'accidental' changes in quantities, due to the limitations of finite-precision arithmetic. Nevertheless, both results and analysis strategies may be of interest in the present case. In particular, perturbations to the inverse operator are generally found to be more pernicious than those in the forward calculation, especially when the linear system is ill-conditioned (e.g. Grcar 2010).

For illustrative purposes, we remark that a natural choice for $\mathcal{S}$ is an operator built using the partial derivatives of $\mathbf{s}(\mathbf{m})$ with respect to the model parameters: in the simplest form,

$$\mathcal{S}_{\mathbf{m_i}}(\mathbf{x}) = \mathbf{A}\mathbf{x} \quad (4a)$$

for any appropriately dimensioned vector $\mathbf{x}$, and where $A_{ij} = \partial s_i/\partial m_j|_{\mathbf{m}=\mathbf{m_i}}$ (sometimes referred to as the Jacobian or Fréchet derivatives matrix). Typically, this does not have an exact inverse, as $\mathbf{A}$ is not square, and may not be full rank. A variety of choices for the inverse operator are therefore possible, but one common choice is the 'least-squares' solution (e.g. Menke 1989), which can be expressed

$$\mathcal{S}_{\mathbf{m_i}}^{-g}(\mathbf{x}) = (\mathbf{A}^\mathrm{T}\mathbf{A} + \mathbf{D})^{-1}\mathbf{A}^\mathrm{T}\mathbf{x} \quad (4b)$$

where $\mathbf{D}$ is some regularization matrix; one common choice is to employ 'Tikhonov regularization', in which case $\mathbf{D} = \epsilon\mathbf{I}$ for some appropriately chosen value of ϵ. We note that a more complete formulation of least squares in the presence of Gaussian uncertainties has been given by Tarantola & Valette (1982); however, this cannot in general be expressed in the form of a linear operator. We will return to this in Section 3.3; nevertheless, this paper is concerned with the behaviour of eq. (3) in general, rather than on any specific choices for the various quantities involved.

2.1 The null spaces

Any linear operator, $\mathcal{L}$, which maps between two vector spaces, can be represented as a matrix, $\mathbf{L}$. In this paper, we are concerned with mappings between an M-dimensional model space and an N-dimensional data space; typically, $N \gg M$. If $\mathcal{L}$ is a forward mapping (i.e. it maps from models into data), then $\mathbf{L}$ will be an $N \times M$

matrix. Every matrix can be characterized by its four 'fundamental subspaces'. We suppose that $\mathbf{L}$ has L linearly independent columns ($L \leq M$; we will assume that $M \leq N$), and—equivalently—L linearly independent rows. These can be regarded as bases for L-dimensional vector spaces, known as the 'column space' and the 'row space' of $\mathbf{L}$, respectively. For any model $\mathbf{m}$, $\mathbf{Lm}$ must lie within the column space of $\mathbf{L}$; similarly, $\mathbf{L^T d}$ will lie in the row space of $\mathbf{L}$ for any $\mathbf{d}$. The column space may also be referred to as the 'image' of the operator $\mathcal{L}$, while the row space is its 'coimage'.

The left null space of $\mathbf{L}$ (or, equivalently, the cokernel of $\mathcal{L}$), is the $(N - L)$-dimensional space representing the portion of the data space that cannot be accessed using $\mathbf{L}$; if $\mathbf{d}$ lies in the left null space, then $\mathbf{L^T d = 0}$. The right null space (or kernel of $\mathcal{L}$) is the set of all models that lie outside the row space of $\mathbf{L}$, and consequently map to zero: if $\mathbf{m}$ lies in the right null space, $\mathbf{Lm = 0}$. Clearly, the right null space has dimension $(M - L)$. The singular value decomposition (SVD) provides a mechanism for accessing these spaces: we can write

$$\mathbf{L} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_0 \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_0^T \end{pmatrix} \tag{5}$$

where $\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_0 \end{pmatrix}$ and $\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_0 \end{pmatrix}$ are unitary matrices, whose columns are referred to as the left and right singular vectors, respectively, and where $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_L)$ with $\sigma_i \geq \sigma_{i+1}$. Then, $\mathbf{U}_1$ consists of L unit vectors spanning the column space of $\mathbf{L}$, while $\mathbf{U}_0$ spans the left null space. Similarly, the L columns of $\mathbf{V}_1$ span the row space of $\mathbf{L}$ and $\mathbf{V}_0$ defines the right null space. Various algorithms exist for computing the SVD efficiently (see e.g. Golub & van Loan 2012); note that in the case where $L = M = N$, the SVD is closely related to matrix diagonalization.

Similarly, the inverse operator $\mathcal{L}^{-g}$ has its own image, coimage, kernel and cokernel. In many cases, the inverse will be constructed so that its image is equivalent to, or a subset of, the coimage of $\mathcal{L}$. However, again, we need not make that assertion here. For simplicity, in the remainder of this paper we shall largely avoid the terms (co-)image and (co-)kernel, and instead refer to each operator as having two image and two null spaces, one each on the data side and the model side. This may not appeal to mathematicians, but in our experience is a common usage within the geophysics community. We do not believe that any confusion should ensue.

2.2 The 'resolution' operator

We can consider the effect of applying eq. (3) to a synthetic data set, corresponding to some known model, $\mathbf{m}_{\mathrm{true}}$. Replacing $\mathbf{d}$ with $\mathbf{s}(\mathbf{m}_{\mathrm{true}})$, and making use of eq. (1), a single model update starting from $\mathbf{m_i}$ will result in a recovered model

$$\mathbf{m}_{\mathrm{rec}} = \mathbf{m_i} + \mathcal{S}_{\mathbf{m_i}}^{-g} \mathcal{S}_{\mathbf{m_i}} (\mathbf{m}_{\mathrm{true}} - \mathbf{m_i}) + \mathcal{S}_{\mathbf{m_i}}^{-g} \mathbf{b}_{\mathcal{S}}(\mathbf{m_i}, \mathbf{m}_{\mathrm{true}}). \tag{6}$$

The compound operator $\mathcal{R}_{\mathbf{m_i}}^{\mathcal{S}} = \mathcal{S}_{\mathbf{m_i}}^{-g} \mathcal{S}_{\mathbf{m_i}}$ is often referred to as the 'resolution' operator, as it provides information on how well a given model can be recovered, or resolved. If $\mathbf{m_i}$ and $\mathbf{m}_{\mathrm{true}}$ are sufficiently close and the operator $\mathcal{S}$ provides a good approximation to the local behaviour of $\mathbf{s}(\mathbf{m})$ in the sense of eq. (2), then the term involving $\mathbf{b}_{\mathcal{S}}$ will be negligible. Then, it can be seen that $\mathbf{m}_{\mathrm{rec}}$ is a version of $\mathbf{m}_{\mathrm{true}}$ subject to any distortions imposed by the imaging process. Unless $\mathcal{S}^{-g}$ is an exact inverse of $\mathcal{S}$, a perfect recovery is not always possible. In particular, we see that models can only be recovered up to an additive null space component: any part of $\mathbf{m}_{\mathrm{true}}$ that lies in the null space of $\mathcal{S}_{\mathbf{m_i}}$ can never be resolved. It is important to take this into account during interpretation and analysis of results

(e.g. de Wit *et al.* 2012). Of course, in practice it is impossible to properly assess whether $\mathbf{m_i}$ is sufficiently close to $\mathbf{m}_{\mathrm{true}}$ for the $\mathbf{b}_{\mathcal{S}}$ term to be neglected, but an assumption that this is true underpins any linearized approach.

If the iterative algorithm in eq. (3) has converged, we must have $\mathbf{m_i} = \mathbf{m}_{i+1} = \mathbf{m}_{\infty}$. In this limit, we see that eq. (6) gives

$$\mathcal{R}_{\mathbf{m_i}}^{\mathcal{S}} (\mathbf{m}_{\infty} - \mathbf{m}_{\mathrm{true}}) = \mathcal{S}_{\mathbf{m_i}}^{-g} \mathbf{b}_{\mathcal{S}} (\mathbf{m}_{\infty}, \mathbf{m}_{\mathrm{true}}). \tag{7}$$

In the case where $\mathbf{b}_{\mathcal{S}}$ is negligible, we see that convergence implies that any difference between $\mathbf{m}_{\mathrm{true}}$ and $\mathbf{m}_{\infty}$—that is, between reality and the recovered image—lies in the null space of the resolution operator. This implies that the model is 'accurate' in so far as it can be resolved by the inversion algorithm. If $\mathbf{b}_{\mathcal{S}}$ is not negligible, some part of the difference will lie within the image space of the resolution operator, and the recovered model contains systematic error. Note that the kernel of $\mathcal{R}^{\mathcal{S}}$ must contain the null space of $\mathcal{S}$, but may be larger; similarly, the image space of $\mathcal{R}^{\mathcal{S}}$ must be a subset of, or equal to, the image space of $\mathcal{S}^{-g}$. In practice, inversion algorithms may not be continued until complete convergence is obtained (i.e. where the model ceases to change between successive iterations): it is common to terminate the inversion procedure once a model that is deemed 'sufficiently good' has been obtained. In such a case, recovered models may deviate from the 'true' model outside the null space of the resolution operator.

2.3 The relative evolution of two linearized inversions

For a given data set, we can compare the results obtained by performing imaging using two separate inversion algorithms. One relies on a forward model $\mathbf{s}(\mathbf{m})$ and an inverse operator $\mathcal{S}^{-g}$, as in eq. (3); the other has the same general form, but uses a different forward model, $\mathbf{f}(\mathbf{p})$, and inverse operator $\mathcal{F}^{-g}$. We assume that in both cases, the model space is parametrized in the same way. We place no restrictions upon the manner in which the operators may differ, but one might envisage comparing two alternative computational implementations of the forward model, perhaps based on different assumptions; or comparing different schemes for regularizing or smoothing the inverse operator. More abstractly, one might wish to compare idealized, 'perfect' forward and inverse operators with those that can be obtained computationally, to explore how results might be affected by approximations, discretization errors, numerical effects and so forth. We discuss some of these possibilities in Section 3.2, below.

However the algorithmic differences arise, we can write a comparison between the solutions at the $(i + 1)$th iteration as

$$\mathbf{\Delta}_{i+1} = \mathbf{\Delta}_i + \left(\mathcal{F}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{m_i}}^{-g} \right) \mathbf{d} + \mathcal{S}_{\mathbf{m_i}}^{-g} \mathbf{s}(\mathbf{m_i}) - \mathcal{F}_{\mathbf{p_i}}^{-g} \mathbf{f}(\mathbf{p_i}) \tag{8a}$$

where we have defined $\mathbf{\Delta}_i = \mathbf{p_i} - \mathbf{m_i}$. Application of eq. (1), or its counterpart for $\mathbf{f}(\mathbf{p})$, allows us to also express this in the alternate forms

$$\mathbf{\Delta}_{i+1} = \left(\mathcal{I} - \mathcal{S}_{\mathbf{m_i}}^{-g} \mathcal{S}_{\mathbf{m_i}} \right) \mathbf{\Delta}_i + \left(\mathcal{F}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{m_i}}^{-g} \right) \mathbf{d} + \mathcal{S}_{\mathbf{m_i}}^{-g} \mathbf{s}(\mathbf{p_i})$$
$$- \mathcal{F}_{\mathbf{p_i}}^{-g} \mathbf{f}(\mathbf{p_i}) - \mathcal{S}_{\mathbf{m_i}}^{-g} \mathbf{b}_{\mathcal{S}} (\mathbf{m_i}, \mathbf{p_i}) \tag{8b}$$

and

$$\mathbf{\Delta}_{i+1} = \left(\mathcal{I} - \mathcal{F}_{\mathbf{p_i}}^{-g} \mathcal{F}_{\mathbf{p_i}} \right) \mathbf{\Delta}_i + \left(\mathcal{F}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{m_i}}^{-g} \right) \mathbf{d} + \mathcal{S}_{\mathbf{m_i}}^{-g} \mathbf{s}(\mathbf{m_i})$$
$$- \mathcal{F}_{\mathbf{p_i}}^{-g} \mathbf{f}(\mathbf{m_i}) + \mathcal{F}_{\mathbf{p_i}}^{-g} \mathbf{b}_{\mathcal{F}} (\mathbf{p_i}, \mathbf{m_i}) \tag{8c}$$

where $\mathcal{I}$ is used to denote the identity operator, and where $\mathbf{b}_{\mathcal{F}}$ is defined in analogy with $\mathbf{b}_{\mathcal{S}}$. The $\mathbf{b}_{\mathcal{S}}$ and $\mathbf{b}_{\mathcal{F}}$ terms arise because in general, the forward problem is known to be non-linear; nevertheless, the inversion algorithms themselves are linearized, as discussed

above. Note that an iteration-by-iteration comparison may not always lend itself to meaningful interpretation: different algorithms may follow very different paths through model space, yet converge to similar points. In such cases, while eq. (8a) remains mathematically well defined, insights into image quality will likely come from consideration of the position at convergence (see e.g. Section 3.1.4). On the other hand, where differences between algorithms are more subtle, the iteration-by-iteration comparison may prove instructive.

3 ANALYSIS AND DISCUSSION

To gain some insight into where differences between two images arise, we can identify a number of different effects contributing to, for example, eq. (8b), which can be expressed in the form

$$\mathbf{\Delta_{i+1}} = \mathbf{\Delta_{i+1}^{res}} + \mathbf{\Delta_{i+1}^{path}} + \mathbf{\Delta_{i+1}^{fwd}} + \mathbf{\Delta_{i+1}^{inv}} + \mathbf{\Delta_{i+1}^{cross}}. \tag{9a}$$

The first term,

$$\mathbf{\Delta_{i+1}^{res}} = \left(\mathcal{I} - \mathcal{S}_{\mathbf{m_i}}^{-g}\mathcal{S}_{\mathbf{m_i}}\right)\mathbf{\Delta_i} - \mathcal{S}_{\mathbf{m_i}}^{-g}\mathbf{b}_{\mathcal{S}}\left(\mathbf{m_i}, \mathbf{p_i}\right) \tag{9b}$$

arises from limitations in the resolution of the imaging process. A comparison with eq. (6) reveals that this term represents the part of $\mathbf{\Delta}_i$ that cannot be accessed by the imaging framework. The second term,

$$\mathbf{\Delta_{i+1}^{path}} = \left(\mathcal{S}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{m_i}}^{-g}\right)\left(\mathbf{d} - \mathbf{s}(\mathbf{p_i})\right), \tag{9c}$$

represents effects that arise because each inversion follows a different path through model space. The third term depends solely on the difference between the forward models,

$$\mathbf{\Delta_{i+1}^{fwd}} = \mathcal{S}_{\mathbf{p_i}}^{-g}\left(\mathbf{s}(\mathbf{p_i}) - \mathbf{f}(\mathbf{p_i})\right), \tag{9d}$$

while the fourth depends solely on the difference between the inverse operators,

$$\mathbf{\Delta_{i+1}^{inv}} = \left(\mathcal{F}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{p_i}}^{-g}\right)\left(\mathbf{d} - \mathbf{s}(\mathbf{p_i})\right). \tag{9e}$$

Finally,

$$\mathbf{\Delta_{i+1}^{cross}} = \left(\mathcal{F}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{p_i}}^{-g}\right)\left(\mathbf{s}(\mathbf{p_i}) - \mathbf{f}(\mathbf{p_i})\right) \tag{9f}$$

represents the interaction between the differences in forward and inverse operators. Clearly this could be absorbed into $\mathbf{\Delta_{i+1}^{fwd}}$ or $\mathbf{\Delta_{i+1}^{inv}}$, but it may assist interpretation to maintain it as a separate term, rather than choosing one or the other of these options.

3.1 Designing an inversion scheme

The effects of any choice made in setting up an inversion scheme can be analysed through eq. (9). It is important to recognize that the relative significance of the various terms in this equation depend on their geometrical relationships, as well as on their magnitudes. By the triangle inequality, we must have

$$\left\|\mathbf{\Delta_{i+1}^{fwd}} + \mathbf{\Delta_{i+1}^{inv}} + \mathbf{\Delta_{i+1}^{cross}}\right\| \leq \left\|\mathbf{\Delta_{i+1}^{fwd}}\right\| + \left\|\mathbf{\Delta_{i+1}^{inv}}\right\| + \left\|\mathbf{\Delta_{i+1}^{cross}}\right\| \tag{10}$$

and thus it need not be the case that two inversion algorithms which rely on a common forward model—or a common inverse operator—will necessarily yield results that are more similar than those of two inversion algorithms which differ in both forward model and inverse operator.

3.1.1 The forward calculation

If two inversion schemes are set up so that at any given point, they use the same inverse operator, but use different forward models for calculating synthetic data, we have $\mathcal{F}^{-g} = \mathcal{S}^{-g}$, and thus

$\mathbf{\Delta_{i+1}^{inv}} = \mathbf{\Delta_{i+1}^{cross}} = \mathbf{0}$. This might entail comparing forward codes built on different numerical integration schemes, or comparing numerical integration with asymptotic methods (for seismological examples, see Cormier 2007). Alternatively, the same forward code might be used in both cases, but with different settings—perhaps to explore the extent to which numerical discretization errors propagate into solutions. If we assume that both inversions start from the same initial model, so that $\mathbf{p_0} = \mathbf{m_0}$, we see from eq. (9) that the driving force for any difference in recovered images must come from $\mathbf{\Delta_{i+1}^{fwd}}$. Resolution and path-dependent effects then govern how these differences accumulate over multiple iterations. Note that these terms contain $\mathcal{S}$ and $\mathbf{b}_{\mathcal{S}}$ which arise from our use of eq. (1) to redefine the point at which $\mathbf{s}$ is evaluated, and this may need to be accounted for if $\mathbf{s}$ is assumed to have any specific form or properties.

From eq. (9d), we see that the differences between forward models are only relevant to the extent that they lie within the data image space of the inverse operator, $\mathcal{S}_{\mathbf{m_i}}^{-g}$; any differences that lie within its data null space will be discarded. If we suppose that $\mathbf{s}$ and $\mathcal{S}^{-g}$ are obtained based on the same—approximate—physical theory, we can ask how changing to a more accurate forward calculation, represented by $\mathbf{f}$, will affect results. We discuss this question in more detail in Section 3.2.3, below, but where approximations are physically motivated, it is quite possible that a significant part of the 'improvement' from $\mathbf{s}$ to $\mathbf{f}$ lie in the null space of $\mathcal{F}^{-g}$. As a trivial illustration, suppose that $\mathbf{s}$ and $\mathbf{S}^{-g}$ are computed for vertical-channel seismic data only, whereas $\mathbf{f}$ is more complete, and includes horizontal motion. This additional information would not, however, be accessible using $\mathcal{F}^{-g}(= \mathcal{S}^{-g})$. The same issue can be viewed from the opposite perspective: if any part of $\mathbf{s}$ can be shown to lie in the null space of $\mathcal{S}^{-g}$, then this can be omitted without affecting results. In some circumstances, this could offer a route to accelerating the forward calculation—and thus, the imaging process as a whole—without degradation in image quality.

3.1.2 The inverse operator

If, instead, the forward models are identical (so that $\mathbf{f} = \mathbf{s}$) but the inverse operators are allowed to differ, we have $\mathbf{\Delta_{i+1}^{fwd}} = \mathbf{\Delta_{i+1}^{cross}} = \mathbf{0}$. Now, any differences between images must be driven by $\mathbf{\Delta_{i+1}^{inv}}$. One can envisage comparing a simple gradient-descent inversion scheme with one that attempts to accelerate convergence by incorporating second-order information, such as a (quasi-)Newton method (e.g. Pratt *et al.* 1998). Similarly, the effects of changes to how data are filtered and processed can also be explored by regarding these as modifications to the inverse operator—see Section 3.2.2. Regardless of how the differences arise, it is important to recognize that $\mathcal{F}^{-g}$ and $\mathcal{S}^{-g}$ need not act in the same subspaces, and that—particularly in high-dimensional settings—even small changes to the operators can result in very different null spaces. Thus, the contribution from $\mathbf{\Delta_{i+1}^{inv}}$ may appear to scale unpredictably with the apparent 'size' of the change in operator.

In the previous section, considering the forward calculation, we have noted that knowledge of the geometric relationships between quantities may enable computational efficiencies. However, this is not generally possible for the inverse operator: $\mathbf{\Delta_{i+1}^{inv}}$ depends upon the data, which cannot usually be known to lie in any particular subspace. Thus, any change in the operator has the potential to affect results. Within our framework, the definition of the inverse operator (see eq. 3) is quite broad, fully encapsulating the connection between observed data and recovered model. Thus—for example—choices regarding data selection and processing, or model parametrization,

are most naturally considered in terms of transformations acting upon this operator. We discuss this point in more detail in Section 3.2, below.

3.1.3 The cross term

The final term, $\boldsymbol{\Delta}_{i+1}^{\mathrm{cross}}$, has a somewhat counter-intuitive form, with the inverse operator from one algorithm acting on the forward operator from the other algorithm. However, interpretation is straightforward if one views eq. (9) as expressing differences in models in terms of the differences in the operators used in their construction. Then, the cross term simply represents the interaction between differences in inverse operators and differences in forward calculations. Plainly, it is only relevant in comparisons where both aspects differ.

If this is the case, $\boldsymbol{\Delta}_{i+1}^{\mathrm{cross}}$ must be considered in conjunction with the effects described in Sections 3.1.1 and 3.1.2. When all three terms are considered together, some simplification is possible, such that

$$\boldsymbol{\Delta}_{i+1}^{\mathrm{fwd}} + \boldsymbol{\Delta}_{i+1}^{\mathrm{inv}} + \boldsymbol{\Delta}_{i+1}^{\mathrm{cross}} = \mathcal{F}_{\mathbf{p_i}}^{-g}\big(\mathbf{d} - \mathbf{f}(\mathbf{p_i})\big) - \mathcal{S}_{\mathbf{p_i}}^{-g}\big(\mathbf{d} - \mathbf{s}(\mathbf{p_i})\big), \quad (11)$$

and this, in turn, will partially cancel with $\boldsymbol{\Delta}_{i+1}^{\mathrm{path}}$. In some cases, combining terms in this manner will simplify the analysis. As has already been observed, there is no reason to assume that the effect of these terms combined will necessarily be less than either $\boldsymbol{\Delta}_{i+1}^{\mathrm{fwd}}$ or $\boldsymbol{\Delta}_{i+1}^{\mathrm{inv}}$ individually. It is quite possible for an inversion scheme that uses approximate methods for obtaining both synthetic data and the inverse operator to give more accurate results than one that computes one of these quantities exactly, and any argument to the contrary must be based upon the specific details of the approximations made.

3.1.4 The position at convergence

Implicit in the use of an iterative algorithm, eq. (3), is the assumption that if sufficient iterations are performed, the system will converge to some particular model: we denote this by $\mathbf{m}_\infty$, as in Section 2.2. Within our comparison of two inversion algorithms, once both have converged we must have $\boldsymbol{\Delta}_i = \boldsymbol{\Delta}_{i+1} = \boldsymbol{\Delta}_\infty$. Thus, again considering eq. (8b), we have

$$\mathcal{S}_{\mathbf{m}_\infty}^{-g} \mathcal{S}_{\mathbf{m}_\infty} \boldsymbol{\Delta}_\infty + \mathcal{S}_{\mathbf{m}_\infty}^{-g} \mathbf{b}_{\mathcal{S}}\left(\mathbf{m}_\infty, \mathbf{p}_\infty\right)$$
$$= \mathcal{S}_{\mathbf{m}_\infty}^{-g}\big(\mathbf{s}(\mathbf{p}_\infty) - \mathbf{f}(\mathbf{p}_\infty)\big) - \mathcal{S}_{\mathbf{m}_\infty}^{-g}\big(\mathbf{d} - \mathbf{f}(\mathbf{p}_\infty)\big), \quad (12)$$

where we have made use of the condition that, at convergence, $\mathcal{F}_{\mathbf{p}_\infty}^{-g}(\mathbf{d} - \mathbf{f}(\mathbf{p}_\infty)) = 0$. The left-hand side of this expression will be recognized as representing the part of $\boldsymbol{\Delta}_\infty$ that can be resolved by the imaging setup—see Section 2.2. The right-hand side could clearly be simplified, but expressing it as two terms assists our interpretation.

The quantity $\mathbf{d} - \mathbf{f}(\mathbf{p}_\infty)$ represents the residuals obtained, at convergence, from the inversion using $\mathbf{f}$ and $\mathcal{F}^{-g}$. By definition, these lie in the data null space of $\mathcal{F}^{-g}$, and are typically assumed to correspond to observational uncertainties, or 'noise'—although this is only true if the assumption that the forward problem encapsulates all relevant physics is also true. If the (inverse) operator is represented as a matrix, the residuals will be orthogonal to every row of that matrix. However, they need not be orthogonal to the matrix representing $\mathcal{S}^{-g}$. In a high-dimensional data space, even small differences between $\mathcal{F}^{-g}$ and $\mathcal{S}^{-g}$ can result in the $\mathcal{S}_{\mathbf{m}_\infty}^{-g}\big(\mathbf{d} - \mathbf{f}(\mathbf{p}_\infty)\big)$ term in eq. (12) becoming significant, particularly where the residuals are relatively large. Thus, even if two algorithms share similar forward codes (so that $\mathbf{s}(\mathbf{p}_\infty) - \mathbf{f}(\mathbf{p}_\infty)$ is small), different images may result.

3.2 Practical applications: towards a better understanding of geophysical images

Having described how the results of two imaging processes differ, it is natural to ask which gives 'better' results. This is not a straightforward question to answer, not least because quantifying image quality is challenging. We reiterate our comments in Valentine & Trampert (2012): a distinction must be drawn between 'predictively accurate' images and 'geologically accurate' images. The model that faithfully describes the true state of a system need not be the model that most accurately explains observations when a given method for performing the forward calculation is used. If the forward method is inaccurate or incomplete in any respect, model alterations may mitigate these deficiencies. For example, some of the effects of using an inaccurate earth model when computing synthetic seismograms can be accommodated by altering the location of the seismic source (e.g. Dziewonski *et al.* 1981), and many surface wave studies rely on 'crustal corrections' to attempt to account for near-surface structure (e.g. Bozdağ & Trampert 2008). Similarly, the model that gives best results in conjunction with one method for performing the forward calculation need not be optimal with another.

3.2.1 Model parametrization and 'nuisance' parameters

Any finite model parametrization is necessarily incomplete: certain types of features or effects will not be representable. A typical tomographic model might specify *S*-wave velocity in a spherical harmonic basis, complete to angular order *L*: in setting up the inverse problem, we assume that there is no structural variation at scale lengths shorter than that corresponding to *L*, and that other physical properties of the Earth—including *P*-wave velocity, density, attenuation and anisotropy—are either known *a priori*, or correlate directly with the *S*-wave velocity. Of course, these assumptions are not truly valid and can introduce biases in results (see e.g. Trampert & Snieder 1996; Valentine & Woodhouse 2010).

We are now in a position to discuss this effect in a very general sense. Suppose we can divide our model parameters into two independent classes, such that $\mathbf{m} = (\mathbf{m}^{(1)}, \mathbf{m}^{(2)})$. Within one inversion scheme, both are allowed to vary; in the other, the second subset take fixed values, so that $\mathbf{p} = (\mathbf{p}^{(1)}, \mathbf{c})$. Thus, the first class might correspond to those components with angular order $l \leq L_1$, while the second class represents higher degree components, such that $L_1 < l \leq L_2$. One can then explore the effect of neglecting the higher degree structure, by choosing $\mathbf{c} = \mathbf{0}$. Regardless of the comparison being made, we assume that the same forward calculation is used in each case, and write $\mathbf{s}(\mathbf{m}_i) = \mathbf{s}(\mathbf{m}_i^{(1)}, \mathbf{m}_i^{(2)})$ and $\mathbf{f}(\mathbf{m}_i) = \mathbf{s}(\mathbf{m}_i^{(1)}, \mathbf{c})$. The forward operator $\mathcal{S}$ can similarly be partitioned, so that $\mathcal{S}_{\mathbf{m}_i} = (\mathcal{S}_{\mathbf{m}_i}^{(1)}, \mathcal{S}_{\mathbf{m}_i}^{(2)})$, while its counterpart for $\mathbf{f}$ must be $\mathcal{F}_{\mathbf{p}_i} = (\mathcal{S}_{\mathbf{p}_i}^{(1)}, 0)$. In practice, this might be achieved by using the same computational implementation in each case: typically, codes that rely on spherical harmonic model parametrizations allow the user to specify the maximum angular order to be considered.

We assume that the inverse operator for $\mathcal{F}$ is constructed to provide information only on those parameters that are allowed to vary, so that $\mathcal{F}_{\mathbf{p}_i}^{-g} = (\mathcal{S}_{\mathbf{p}_i}^{(1)\,-g}, 0)$. From eq. (8c), we therefore see that

$$\big(\boldsymbol{\Delta}_{i+1}^{(1)}, \boldsymbol{\Delta}_{i+1}^{(2)}\big) = \big((\mathcal{I} - \mathcal{S}_{\mathbf{p}_i}^{(1)\,-g}\mathcal{S}_{\mathbf{p}_i}^{(1)})\boldsymbol{\Delta}_i^{(1)} + \mathcal{S}_{\mathbf{p}_i}^{(1)\,-g}\big(\mathbf{d} - \mathbf{s}(\mathbf{p}_i^{(1)}, \mathbf{c})\big)$$
$$+ \mathcal{S}_{\mathbf{p}_i}^{(1)\,-g}\mathbf{b}_{\mathcal{S}}\big(\mathbf{p}_i^{(1)}, \mathbf{c}; \mathbf{m}_i^{(1)}, \mathbf{c}\big), \boldsymbol{\Delta}_i^{(2)}\big)$$
$$- \big(\mathcal{S}_{\mathbf{m}_i}^{(1)}, \mathcal{S}_{\mathbf{m}_i}^{(2)}\big)^{-g}\big(\mathbf{d} - \mathbf{s}(\mathbf{m}_i^{(1)}, \mathbf{m}_i^{(2)})\big) \quad (13a)$$

The relative evolution of the two images therefore depends upon the construction of $(\mathcal{S}_{\mathbf{m_i}}^{(1)}, \mathcal{S}_{\mathbf{m_i}}^{(2)})^{-g}$. If the inverse is constructed in such a way that the distinction between the two sets of model parameters is maintained, allowing us to write $(\mathcal{S}_{\mathbf{m_i}}^{(1)}, \mathcal{S}_{\mathbf{m_i}}^{(2)})^{-g} = (\mathcal{S}_{\mathbf{m_i}}^{(1)^{-g}}, \mathcal{S}_{\mathbf{m_i}}^{(2)^{-g}})$, then eq. (13a) may be expressed

$$\boldsymbol{\Delta}_{i+1}^{(1)} = \left(\mathcal{I} - \mathcal{S}_{\mathbf{p_i}}^{(1)^{-g}} \mathcal{S}_{\mathbf{p_i}}^{(1)}\right) \boldsymbol{\Delta}_{i}^{(1)} + \left(\mathcal{S}_{\mathbf{p_i}}^{(1)^{-g}} - \mathcal{S}_{\mathbf{m_i}}^{(1)^{-g}}\right) \mathbf{d}$$

$$+ \mathcal{S}_{\mathbf{m_i}}^{(1)^{-g}} \mathbf{s}\left(\mathbf{m_i}^{(1)}, \mathbf{m_i}^{(2)}\right) - \mathcal{S}_{\mathbf{p_i}}^{(1)^{-g}} \mathbf{s}\left(\mathbf{p_i}^{(1)}, \mathbf{c}\right)$$

$$+ \mathcal{S}_{\mathbf{p_i}}^{(1)^{-g}} \mathbf{b}_{\mathcal{S}}\left(\mathbf{p_i}^{(1)}, \mathbf{c}; \mathbf{m_i}^{(1)}, \mathbf{c}\right)$$

$$\boldsymbol{\Delta}_{i+1}^{(2)} = \boldsymbol{\Delta}_{i}^{(2)} - \mathcal{S}_{\mathbf{m_i}}^{(2)^{-g}}\left(\mathbf{d} - \mathbf{s}\left(\mathbf{m_i}^{(1)}, \mathbf{m_i}^{(2)}\right)\right). \tag{13b}$$

If $\mathbf{p_i}^{(1)} = \mathbf{m_i}^{(1)}$, so that $\boldsymbol{\Delta}_i^{(1)} = 0$, and the linear operator $\mathcal{S}$ is accurate, in the sense that $\mathbf{b}_{\mathcal{S}}(\mathbf{m_i}, \mathbf{m_i}) = 0$, then

$$\boldsymbol{\Delta}_{i+1}^{(1)} = \mathcal{S}_{\mathbf{m_i}}^{(1)^{-g}}\left(\mathbf{s}\left(\mathbf{m_i}^{(1)}, \mathbf{m_i}^{(2)}\right) - \mathbf{s}\left(\mathbf{m_i}^{(1)}, \mathbf{c}\right)\right)$$

$$= -\mathcal{S}_{\mathbf{m_i}}^{(1)^{-g}}\left(\mathcal{S}_{\mathbf{m_i}}^{(2)} \boldsymbol{\Delta}_i^{(2)} + \mathbf{b}_{\mathcal{S}}\left(\mathbf{m_i}^{(1)}, \mathbf{m_i}^{(2)}; \mathbf{m_i}^{(1)}, \mathbf{c}\right)\right)$$

$$\boldsymbol{\Delta}_{i+1}^{(2)} = \boldsymbol{\Delta}_i^{(2)} - \mathcal{S}_{\mathbf{m_i}}^{(2)^{-g}}\left(\mathbf{d} - \mathbf{s}\left(\mathbf{m_i}^{(1)}, \mathbf{m_i}^{(2)}\right)\right) \tag{13c}$$

where we have made use of eq. (1). Thus, we see that for the two inversion algorithms to agree over the first subset of parameters, $\mathcal{S}_{\mathbf{m_i}}^{(2)} \boldsymbol{\Delta}_i^{(2)} + \mathbf{b}_{\mathcal{S}}(\mathbf{m_i}^{(1)}, \mathbf{m_i}^{(2)}; \mathbf{m_i}^{(1)}, \mathbf{c})$ must lie in the data null space of $\mathcal{S}_{\mathbf{m_i}}^{(1)^{-g}}$. In general, this will arise if the separation between the two sets of model parameters also corresponds to a separation in the data space—that is, if the data image spaces of $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ do not overlap. Our requirement that the inverse is constructed from each subspace separately follows naturally from this condition. Thus, under these circumstances, provided the two inversions share a common starting point, they will never diverge for the first subset of parameters; obviously, they will disagree for the second subset. However, a complete data-space separation is rare in non-trivial examples.

In the more common case, where the two subsets of model parameters give rise to effects that overlap in the data space, there will generally be coupling between $\boldsymbol{\Delta}_i^{(2)}$ and $\boldsymbol{\Delta}_{i+1}^{(1)}$. Thus, fixing some parameters to erroneous values will result in a bias appearing in the remaining parameters, in order to compensate for this. The model with fixed parameters will be predictively accurate (in the sense defined at the start of Section 3.2) and may give better results than a geologically accurate model used in conjunction with those fixed parameters. Unfortunately, it is almost always infeasible to invert for all possible physical parameters simultaneously and fixed parameters must be used (either explicitly, or implicitly). Thus, some bias in results will usually be present, and this must be borne in mind during interpretation.

3.2.2 Data-space transformations

Invariably, some form of pre-processing is carried out upon data prior to imaging, and this can also be explored within our framework, comparing results between processed and unprocessed data sets. In many cases, these operations can best be represented as a transformation of the imaging operator: typically the forward calculation is not directly altered (albeit that some methods of computation may permit certain efficiencies—for example, normal-mode-based calculations might omit frequencies that would be lost during filtering). We therefore identify $\mathcal{S}$ and $\mathbf{s}$ with the unpreprocessed data, while $\mathcal{F}^{-g} = (\mathcal{P}\mathcal{S})^{-g}\mathcal{P}$ for some processing operator $\mathcal{P}$, and

$\mathbf{f} = \mathbf{s}$. Thus, $\boldsymbol{\Delta}_{i+1}^{\mathrm{fwd}} = \boldsymbol{\Delta}_{i+1}^{\mathrm{cross}} = 0$, and any differences in results arise primarily from

$$\boldsymbol{\Delta}_{i+1}^{\mathrm{inv}} = \left((\mathcal{P}\mathcal{S}_{\mathbf{m_i}})^{-g}\mathcal{P} - \mathcal{S}_{\mathbf{m_i}}^{-g}\right)\left(\mathbf{d} - \mathbf{s}(\mathbf{m_i})\right). \tag{14}$$

In general, if the two images are to be identical, we must have $(\mathcal{P}\mathcal{S})^{-g}\mathcal{P} = \mathcal{S}^{-g}$. This will trivially be the case if the transformation $\mathcal{P}$ acts only on data components that lie in the data null space of $\mathcal{S}$. More interestingly, we can consider partitioning the data space into two parts, such that $\mathbf{d} = (\mathbf{d}^{(1)}, \mathbf{d}^{(2)})$ and write $\mathcal{S} = \mathcal{S}^{(1)} + \mathcal{S}^{(2)}$, where $\mathbf{d}^{(1)}$ must lie in the image space of $\mathcal{S}^{(1)}$ and $\mathbf{d}^{(2)}$ in the image space of $\mathcal{S}^{(2)}$ (note that this partitioning is unrelated to that discussed in Section 3.2.1). Suppose, further, that the effect of $\mathcal{P}$ is to discard any data-space components that lie in the second subspace, that is, $\mathcal{P}\mathbf{d} = (\mathbf{d}^{(1)}, \mathbf{0})$. In this case, $\mathcal{P}\mathcal{S} = \mathcal{S}^{(1)}$, and therefore $\mathcal{F}^{-g} = \mathcal{S}^{(1)^{-g}}\mathcal{P}$. Hence, we obtain

$$\boldsymbol{\Delta}_{i+1}^{\mathrm{inv}} = \left(\mathcal{S}_{\mathbf{m_i}}^{(1)}\right)^{-g}\left(\mathbf{d}^{(1)} - \mathbf{s}^{(1)}(\mathbf{m_i})\right)$$

$$- \left(\mathcal{S}_{\mathbf{m_i}}^{(1)} + \mathcal{S}_{\mathbf{m_i}}^{(2)}\right)^{-g}\left(\mathbf{d} - \mathbf{s}(\mathbf{m_i})\right). \tag{15a}$$

Again, if the inverse operator maintains the separation between spaces, such that $\left(\mathcal{S}_{\mathbf{m_i}}^{(1)} + \mathcal{S}_{\mathbf{m_i}}^{(2)}\right)^{-g} = \mathcal{S}_{\mathbf{m_i}}^{(1)^{-g}} + \mathcal{S}_{\mathbf{m_i}}^{(2)^{-g}}$, this expression can be simplified, and we obtain

$$\boldsymbol{\Delta}_{i+1}^{\mathrm{inv}} = \left(\mathcal{S}_{\mathbf{m_i}}^{(2)}\right)^{-g}\left(\mathbf{d}^{(2)} - \mathbf{s}^{(2)}(\mathbf{m_i})\right). \tag{15b}$$

This essentially requires the two operators, $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$, to have distinct coimage spaces in addition to the distinct image spaces (which arise by definition). In this case, the processing operation will not affect our ability to image the model parameters in the coimage space of $\mathcal{S}^{(1)}$. This is essentially the same situation as discussed in the previous section, but approached from the opposite direction. Note that even where $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ act between non-overlapping spaces, it is still possible to construct an inverse operator that does not preserve this separation—for example, through the use of a regularization scheme which does not treat the two operators independently.

In certain cases, it may be possible to show that $(\mathcal{P}\mathcal{S})^{-g} = \mathcal{S}^{-g}\mathcal{P}^{-g}$. This is always true if the inverse is exact, and may also arise under other circumstances. If, additionally, the transformation $\mathcal{P}$ satisfies $\mathcal{P}^{-g}\mathcal{P} = \mathcal{I}$, it will not affect the results of inversion. Note that any transformation that discards information—such as a filtering operation—cannot satisfy this requirement, and will therefore alter the recovered image. However, it is straightforward to demonstrate that transformations such as rescaling of data need not affect results (although in practice, these transformations may have important consequences for the numerical stability of a given computational implementation).

As a final remark in this section: note that in practice, solution of an inverse problem according to eq. (3) requires us to compute $\mathbf{s}$ and $\mathcal{S}^{-g}$. Certain techniques for computing the partial derivatives—such as adjoint methods—are susceptible to numerical noise, and it is normal to attempt to remove this through some form of filtering or smoothing operation. Using $\mathcal{S}$ to denote the theoretical (noise-free) linear operator, and writing $\mathcal{F} = \mathcal{P}(\mathcal{S} + \mathcal{N})$ for noise $\mathcal{N}$ and filter $\mathcal{P}$, we again see that any errors that arise relative to the theoretical image are driven by

$$\boldsymbol{\Delta}_{i+1}^{\mathrm{inv}} = ((\mathcal{P}(\mathcal{S}_{\mathbf{m_i}} + \mathcal{N}))^{-g} - \mathcal{S}_{\mathbf{m_i}}^{-g})(\mathbf{d} - \mathbf{s}(\mathbf{m_i})). \tag{16a}$$

To make more detailed comments, we must make assumptions about the form of various quantities. For illustrative purposes, we consider a simple gradient descent algorithm: if $\mathcal{S}_{\mathbf{m_i}}$ is represented by the matrix $\mathbf{S}$, then $\mathcal{S}_{\mathbf{m_i}}^{-g} = -\alpha \mathbf{S}^{\mathbf{T}}$, for some small α. Thus, assuming $\mathcal{P}$

is a linear operator and can be represented by a matrix $\mathbf{P}$ such that $\mathcal{P}\mathbf{S} = \mathbf{PS}$, we can write

$$\mathbf{\Delta}_{i+1}^{inv} = -\alpha \mathbf{S}^{T}\left(\mathbf{P}^{T} - \mathbf{I}\right)\left(\mathbf{d} - \mathbf{s}(\mathbf{m}_i)\right) - \alpha \mathbf{N}^{T}\mathbf{P}^{T}\left(\mathbf{d} - \mathbf{s}(\mathbf{m}_i)\right).\quad(16\text{b})$$

It is clear that there is a tradeoff between the elimination of noise, on one hand, and the loss of useful information, on the other. This cannot be avoided unless $\mathbf{N}$ can be shown to lie wholly in a different data subspace from $\mathbf{S}$.

3.2.3 'Hybrid' schemes

For many geophysical systems, we have a fairly complete understanding of the physical processes that give rise to observable phenomena. Thus, we have the ability to compute synthetic data in a theoretically complete manner—for example, fully numerical techniques (e.g. Komatitsch *et al.* 2002) allow simulation of seismic wavefields with almost arbitrary accuracy. Similarly, exact partial derivatives may be obtained, and it is clearly desirable to use the most accurate techniques available when imaging a geophysical system. However, these methods are invariably computationally demanding, and in many situations, their use may not currently be feasible.

A natural question is therefore whether a 'hybrid' imaging scheme may usefully be implemented, which combines certain calculations performed using 'fully accurate' techniques with others performed using an approximate—and therefore computationally cheaper—method. In particular, the use of accurate synthetics in conjunction with an inverse operator computed from approximate partial derivatives has sometimes been proposed. Notable recent implementations of this concept include the work of Lekić & Romanowicz (2011) and French *et al.* (2013) (although these do not rely on an inversion algorithm that is strictly linear; see Section 3.3). We are now in a position to consider the effects of such an approach, which may be seen from two perspectives: as an 'improvement' in the forward model used, relative to a 'fully approximate' inversion; or as a degradation in the inverse operator relative to a 'fully accurate' approach.

Fundamentally, this form of hybrid approach involves using an inverse operator that is not directly related to the forward calculation. Within the framework established in Section 2, this can be approached from two directions. From one viewpoint, the forward operator in eq. (1), $\mathcal{S}_{\mathbf{m}}$, may be regarded as remaining accurate (in the sense that $\mathbf{b}_{\mathcal{S}}(\mathbf{m}, \mathbf{m}) = 0$) but the inverse operator $\mathcal{S}_{\mathbf{m}}^{-g}$ is not constructed as a direct relative of $\mathcal{S}_{\mathbf{m}}$. In the alternative view, $\mathcal{S}_{\mathbf{m}}^{-g}$ is a 'good' inverse for $\mathcal{S}_{\mathbf{m}}$, but the forward operator does not accurately describe the behaviour of $\mathbf{s}(\mathbf{m})$ and $\mathbf{b}_{\mathcal{S}}(\mathbf{m}, \mathbf{m}) \neq 0$.

We begin by comparing the results of a hybrid scheme to those of fully approximate inversion. The clearest interpretation within the framework of eq. (9) comes from associating $\mathbf{m}$ with the fully approximate case, and $\mathbf{p}$ with the hybrid model. For clarity, we use $\mathbf{h}$ to represent a hybrid model vector, and make the transformation $\mathbf{p} \rightarrow \mathbf{h}$. Similarly, we specify the fully approximate model by $\mathbf{m} \rightarrow \mathbf{a}$. We use $\mathbf{t}(\mathbf{m})$ to represent the exact ('true') forward calculation, and thus transform $\mathbf{f} \rightarrow \mathbf{t}$. We continue to use $\mathbf{s}$ and $\mathcal{S}$ to denote a generic approximate method and its corresponding linear operator. The non-zero terms from eq. (9) may then be expressed

$$\mathbf{\Delta}_{i+1}^{res} = \left(\mathcal{I} - \mathcal{S}_{\mathbf{a}_i}^{-g}\mathcal{S}_{\mathbf{a}_i}\right)\mathbf{\Delta}_i - \mathcal{S}_{\mathbf{a}_i}^{-g}\mathbf{b}_{\mathcal{S}}\left(\mathbf{a}_i, \mathbf{h}_i\right)$$

$$\mathbf{\Delta}_{i+1}^{path} = \left(\mathcal{S}_{\mathbf{h}_i}^{-g} - \mathcal{S}_{\mathbf{a}_i}^{-g}\right)\left(\mathbf{d} - \mathbf{s}(\mathbf{h}_i)\right)$$

$$\mathbf{\Delta}_{i+1}^{fwd} = \mathcal{S}_{\mathbf{h}_i}^{-g}\left(\mathbf{s}(\mathbf{h}_i) - \mathbf{t}(\mathbf{h}_i)\right).\quad(17)$$

Any gain that is associated with the switch to using the exact forward code must come from the action of the operator $\mathcal{S}_{\mathbf{h}_i}^{-g}$ upon $\left(\mathbf{s}(\mathbf{h}_i) - \mathbf{t}(\mathbf{h}_i)\right)$.

We can also make the identification in the opposite sense, so that $\mathbf{m}$ is associated with the hybrid case. In this case, $\mathbf{m} \rightarrow \mathbf{h}$ and $\mathbf{s} \rightarrow \mathbf{t}$. The linear operator that describes the behaviour of $\mathbf{t}$ is $\mathcal{T}$, and we transform $\mathcal{S} \rightarrow \mathcal{T}$, but as already discussed, we are using an incorrect linear operator, so $\mathcal{S}^{-g}$ is not transformed.

$$\mathbf{\Delta}_{i+1}^{res} = \left(\mathcal{I} - \mathcal{S}_{\mathbf{h}_i}^{-g}\mathcal{T}_{\mathbf{h}_i}\right)\mathbf{\Delta}_i - \mathcal{S}_{\mathbf{h}_i}^{-g}\mathbf{b}_{\mathcal{S}}\left(\mathbf{h}_i, \mathbf{a}_i\right)$$

$$\mathbf{\Delta}_{i+1}^{path} = \left(\mathcal{S}_{\mathbf{a}_i}^{-g} - \mathcal{S}_{\mathbf{h}_i}^{-g}\right)\left(\mathbf{d} - \mathbf{t}(\mathbf{a}_i)\right)$$

$$\mathbf{\Delta}_{i+1}^{fwd} = \mathcal{S}_{\mathbf{a}_i}^{-g}\left(\mathbf{t}(\mathbf{a}_i) - \mathbf{s}(\mathbf{a}_i)\right).\quad(18)$$

Note that since $\mathbf{\Delta} = \mathbf{p} - \mathbf{m}$, some signs must be changed for eq. (18) to be equivalent to eq. (17). In any case, interpretation is complicated by the fact that all three terms now depend explicitly upon both the exact and the approximate theories.

The value of eq. (18) becomes apparent when we consider another comparison: between hybrid and exact inversions. Now, we identify $\mathbf{m}$ with the hybrid model, so that $\mathbf{m} \rightarrow \mathbf{h}$ and use $\mathbf{p} \rightarrow \mathbf{e}$ to denote the exact model. Thus, $\mathcal{F}^{-g} \rightarrow \mathcal{T}^{-g}$, while $\mathbf{f} \rightarrow \mathbf{t}$ and $\mathbf{s} \rightarrow \mathbf{t}$. The non-zero terms become

$$\mathbf{\Delta}_{i+1}^{res} = \left(\mathcal{I} - \mathcal{S}_{\mathbf{h}_i}^{-g}\mathcal{T}_{\mathbf{h}_i}\right)\mathbf{\Delta}_i - \mathcal{S}_{\mathbf{h}_i}^{-g}\mathbf{b}_{\mathcal{S}}\left(\mathbf{h}_i, \mathbf{e}_i\right)$$

$$\mathbf{\Delta}_{i+1}^{path} = \left(\mathcal{S}_{\mathbf{e}_i}^{-g} - \mathcal{S}_{\mathbf{h}_i}^{-g}\right)\left(\mathbf{d} - \mathbf{t}(\mathbf{e}_i)\right)$$

$$\mathbf{\Delta}_{i+1}^{inv} = \left(\mathcal{T}_{\mathbf{e}_i}^{-g} - \mathcal{S}_{\mathbf{e}_i}^{-g}\right)\left(\mathbf{d} - \mathbf{t}(\mathbf{e}_i)\right)\quad(19)$$

Comparing this to eq. (18), we see that—apart from differing with regard to the points at which the various quantities are evaluated—$\mathbf{\Delta}_{i+1}^{res}$ and $\mathbf{\Delta}_{i+1}^{path}$ are identical in form in the two cases. Absent any argument founded upon the particular characteristics of a given imaging problem, these terms must be assumed to make a similar contribution to $\mathbf{\Delta}_{i+1}$ in each case. Similarly, we do not see that any general statements can be made about the relative magnitudes of $\mathbf{\Delta}_{i+1}^{fwd}$ and $\mathbf{\Delta}_{i+1}^{inv}$—although we note that the latter depends upon the noise content of the data, and is, in principle, unbounded. However, assuming that data are subject to some degree of quality control, it seems difficult to conclude that one term is always, or often, much larger than the other. Thus, there would appear to be no general reason to suppose that hybrid inversion schemes offer a significant improvement over fully approximate inversion. In order to argue that this is true in any specific case, it will be necessary to consider the particular properties of the approximations used.

3.2.4 The Hessian operator

Finally, we briefly discuss the role of Hessian information, which is sometimes discussed as a route toward improving geophysical imaging, especially in the context of 'full-waveform inversion' of seismic data (e.g. Virieux & Operto 2009; Fichtner & Trampert 2011a). Formally, a 'Hessian' is the $N \times N$ matrix $\mathbf{H}$ with components $H_{ij} = \partial^2 f/\partial x_i \partial x_j$, for any scalar function of N variables $f(x_1, x_2, \ldots, x_N)$. In a geophysical context, the term is invariably used to denote the Hessian of some misfit function $\chi(\mathbf{m})$, which the inversion procedure seeks to minimize. In particular, taking $\mathbf{m}_\infty$ to represent an optimal model—for which, by definition, $\partial \chi/\partial x_i = 0$ for all x_i—the Hessian characterizes the leading term in a Taylor expansion of the misfit function,

$$\chi(\mathbf{m}) = \chi(\mathbf{m}_\infty) + \frac{1}{2}\left(\mathbf{m} - \mathbf{m}_\infty\right)\mathbf{H}_{\mathbf{m}_\infty}\left(\mathbf{m} - \mathbf{m}_\infty\right)$$

$$+ \text{ higher order terms}\quad(20)$$

where we have placed a subscript on $\mathbf{H}$ to indicate that the Hessian must be evaluated at a particular point in model space. It is apparent from eq. (20) that the Hessian provides, to a first approximation, information about the degree to which this optimal solution can be constrained (or, equivalently, on the uncertainties and tradeoffs associated with the solution). It is therefore often beneficial to make use of information contained within this Hessian—or an approximation thereto—when constructing the inverse operator $\mathcal{S}^{-g}$ of eq. (3). Thus, for example, the inverse of the Hessian can be used as a pre-conditioner in gradient descent, improving performance in terms of both convergence rate and quality of results (e.g. Operto *et al.* 2013). However, this approach—commonly referred to as Newton's method—remains a linear algorithm for the computation of each model update. It therefore continues to have the form given in eq. (3), and all our analysis still applies.

In order to compute this Hessian, it will typically be necessary to evaluate the second derivative of the synthetic observables with respect to model parameters. Given this capability, it is natural to ask whether it can be used to gain additional understanding of the behaviour of imaging algorithms. Viewing eq. (1) as a Taylor expansion to first order, it is clear that any information available from second derivatives has been subsumed into $\mathbf{b}_\mathcal{S}$ in our analysis. In eq. (9), we see that this appears only in the resolution term, $\mathbf{\Delta}^{\mathrm{res}}$. This is, perhaps, unsurprising given the Hessian's role in describing the extent to which different parameters are coupled. However, it does not appear that explicit incorporation of second-order information would greatly enhance our ability to understand how differences between images can arise.

3.3 Incorporating resolution information

Returning again to eq. (3), we may appreciate a subtlety. At any given iteration, the model update must lie in the image space of $\mathcal{S}_{\mathbf{m_i}}^{-g}$. Typically, this inverse will be constructed from $\mathcal{S}_{\mathbf{m_i}}$, and the model image space of the inverse will be a subset of, or equal to, that of the forward operator. This is a desirable construction, since it implies that the model update lies wholly within the part of model space which is known to affect synthetic data: no part of the update will lie in the model null space. This is, essentially, the principle of 'Occam's razor': we should add nothing to the model that is not positively required to explain observations.

However, there is no reason to suppose that the fundamental subspaces of $\mathcal{S}_{\mathbf{m_i}}$ will remain unchanged over multiple iterations. An update that contained no component in the model null space computed for the first iteration may be substantially in the null space as computed at the final iteration. This is seen in eq. (9b), where $\mathbf{\Delta}_{\mathbf{i+1}}^{\mathrm{res}}$ allows any null space components to accumulate over multiple iterations. There is no positive reason to retain these, which depend entirely on the iteration-by-iteration history of the inversion, and are likely to mislead during interpretation.

The solution is straightforward: discard any part of the existing model that lies in the null space of the current inverse operator, prior to performing the model update. We define some operator, $\mathcal{V}_{\mathbf{m_i}}^\mathcal{S}$, to select only those model components that lie within the model image space of $\mathcal{S}_{\mathbf{m_i}}$, and iterate according to

$$\mathbf{m_{i+1}} = \mathcal{V}_{\mathbf{m_i}}^\mathcal{S} \mathbf{m_i} + \mathcal{S}_{\mathbf{m_i}}^{-g}\left(\mathbf{d} - \mathbf{s}(\mathbf{m_i})\right). \qquad (21)$$

We can compare images produced by two different implementations of this framework, as before. The counterpart to eq. (8a) can be expressed

$$\mathbf{\Delta_{i+1}} = \mathcal{V}_{\mathbf{m_i}}^\mathcal{S} \mathbf{\Delta_i} + \left(\mathcal{V}_{\mathbf{p_i}}^\mathcal{F} - \mathcal{V}_{\mathbf{m_i}}^\mathcal{S}\right)\mathbf{p_i} + \left(\mathcal{F}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{m_i}}^{-g}\right)\mathbf{d}$$
$$+ \mathcal{S}_{\mathbf{m_i}}^{-g}\mathbf{s}(\mathbf{m_i}) - \mathcal{F}_{\mathbf{p_i}}^{-g}\mathbf{f}(\mathbf{p_i}), \qquad (22a)$$

where $\mathcal{V}_{\mathbf{p_i}}^\mathcal{F}$ is the corresponding operator constructed for $\mathcal{F}_{\mathbf{p_i}}$. Similarly, eq. (8b) becomes

$$\mathbf{\Delta_{i+1}} = \left(\mathcal{V}_{\mathbf{m_i}}^\mathcal{S} - \mathcal{S}_{\mathbf{m_i}}^{-g}\mathcal{S}_{\mathbf{m_i}}\right)\mathbf{\Delta_i} + \left(\mathcal{V}_{\mathbf{p_i}}^\mathcal{F} - \mathcal{V}_{\mathbf{m_i}}^\mathcal{S}\right)\mathbf{p_i} + \left(\mathcal{F}_{\mathbf{p_i}}^{-g} - \mathcal{S}_{\mathbf{m_i}}^{-g}\right)\mathbf{d}$$
$$+ \mathcal{S}_{\mathbf{m_i}}^{-g}\mathbf{s}(\mathbf{p_i}) - \mathcal{F}_{\mathbf{p_i}}^{-g}\mathbf{f}(\mathbf{p_i}) - \mathcal{S}_{\mathbf{m_i}}^{-g}\mathbf{b}_\mathcal{S}\left(\mathbf{m_i}, \mathbf{p_i}\right). \qquad (22b)$$

Thus, the only change in eq. (9) is that eq. (9b) should now be written

$$\mathbf{\Delta}_{\mathbf{i+1}}^{\mathrm{res}} = \left(\mathcal{V}_{\mathbf{m_i}}^\mathcal{S} - \mathcal{S}_{\mathbf{m_i}}^{-g}\mathcal{S}_{\mathbf{m_i}}\right)\mathbf{\Delta_i} - \mathcal{S}_{\mathbf{m_i}}^{-g}\mathbf{b}_\mathcal{S}(\mathbf{m_i}, \mathbf{p_i})$$
$$+ \left(\mathcal{V}_{\mathbf{p_i}}^\mathcal{F} - \mathcal{V}_{\mathbf{m_i}}^\mathcal{S}\right)\mathbf{p_i}. \qquad (23)$$

Provided that $\mathcal{S}^{-g}$ is constructed to map solely into the model image space of $\mathcal{S}$, it is possible—and sensible—to choose $\mathcal{V}_{\mathbf{m_i}}^\mathcal{S} = \mathcal{S}_{\mathbf{m_i}}^{-g}\mathcal{S}_{\mathbf{m_i}} = \mathcal{R}_{\mathbf{m_i}}^\mathcal{S}$. This choice arises naturally in the work of Tarantola & Valette (1982), where a complete Bayesian treatment of the least-squares problem is given for the case of Gaussian uncertainties.

3.3.1 The position at convergence

As in Section 3.1.4, we can consider performing sufficient iterations so that both inversion schemes converge. In this case, we can write down a counterpart to eq. (12), based on eq. (22b)

$$\left(\mathcal{I} - \mathcal{V}_{\mathbf{m_\infty}}^\mathcal{S} + \mathcal{S}_{\mathbf{m_\infty}}^{-g}\mathcal{S}_{\mathbf{m_\infty}}\right)\mathbf{\Delta_\infty}$$
$$= \left(\mathcal{V}_{\mathbf{p_\infty}}^\mathcal{F} - \mathcal{V}_{\mathbf{m_\infty}}^\mathcal{S}\right)\mathbf{p_\infty} - \mathcal{S}_{\mathbf{m_\infty}}^{-g}(\mathbf{d} - \mathbf{f}(\mathbf{p_\infty}))$$
$$+ \mathcal{S}_{\mathbf{m_\infty}}^{-g}(\mathbf{s}(\mathbf{p_\infty}) - \mathbf{f}(\mathbf{p_\infty})) - \mathcal{S}_{\mathbf{m_\infty}}^{-g}\mathbf{b}_\mathcal{S}(\mathbf{m_\infty}, \mathbf{p_\infty}). \qquad (24)$$

In the case where $\mathcal{V}_{\mathbf{m_i}}^\mathcal{S} = \mathcal{S}_{\mathbf{m_i}}^{-g}\mathcal{S}_{\mathbf{m_i}}$, the left-hand side reduces to $\mathbf{\Delta_\infty}$; we have moved the term involving $\mathbf{b}_\mathcal{S}$ to the right-hand side for this reason. As before, the orientation of residuals from one inversion scheme relative to the inverse operator of the other remains important, but there is an additional term to account for the fact that the two imaging operators may map into different subspaces.

3.3.2 A caveat: 'hybrid' inversion schemes

When both forward and inverse operators share common image and null spaces, constructing $\mathcal{V}$ is reasonably straightforward. However, this is not necessarily the case for hybrid inversion schemes. Our 'Occam's razor' argument indicates that the correct choice should identify the model image space of the forward operator—that is, the operator that accurately describes the local behaviour of the synthetic calculation. However, the motivation for performing a hybrid inversion is to avoid computing this operator, and thus the correct subspace is not known. In consequence, it is necessary to construct $\mathcal{V}$ based on the image space of the approximate forward operator, instead. Thus, application of eq. (21) results in a model that may contain features that are not necessary to allow the exact forward theory to explain the data, and for which there is no positive evidence. It is also possible for this to give rise to instabilities in the inversion algorithm, since the model is continually 'corrected' in a manner that does not conform to the behaviour of the forward code used. In particular, there is no reason to suppose that partial derivatives calculated using approximations 'point towards' the model that minimizes misfit between accurate synthetics and data. Thus, the algorithm can never converge. In order to compensate for this, it is likely that stronger regularization would be required, and this may undo some of the benefits associated with the hybrid approach.

4 ILLUSTRATION: DETERMINING THE DENSITY DISTRIBUTION IN A VIBRATING STRING

To illustrate the issues discussed in this paper, we consider a simple synthetic imaging problem: determination of the density distribution within a vibrating string. We assume the string to have unit length, and to be under unit tension. At time t, the displacement of the string is described by the function $y(x, t)$, where x represents the position along the string's length ($0 \leq x \leq 1$). The string has non-uniform density, described by $\rho(x)$, and attenuates in some simple manner, represented by a damping coefficient α. We therefore consider the wave equation

$$\frac{\partial^2 y}{\partial x^2} = \rho(x) \frac{\partial^2 y}{\partial t^2} + \alpha \frac{\partial y}{\partial t}, \tag{25}$$

subject to the boundary conditions $y(0, t) = y(1, t) = 0$, and some initial configuration, $y(x, 0) = y_0(x)$. We assume the string is initially at rest, so that $\dot{y}(x, 0) = 0$, where we have adopted the Newtonian 'dot' notation to represent a time derivative.

4.1 Method of solution

For our purposes, it is helpful to adopt a 'normal-mode' approach to solving this equation. Any solution that satisfies the boundary conditions can be written in the form

$$y(x, t) = \sum_{n=1}^{\infty} a_n(t) \sin(n\pi x). \tag{26}$$

Substituting this into eq. (25) and making use of the orthogonality properties of sinusoids, we find that the oscillation of the nth mode is governed by

$$a_n(t) = \frac{-1}{n^2 \pi^2} \left[\alpha \frac{\mathrm{d}a_n}{\mathrm{d}t} + 2 \sum_m \frac{\mathrm{d}^2 a_m}{\mathrm{d}t^2} \right.$$
$$\left. \times \int_0^1 \rho(x) \sin(m\pi x) \sin(n\pi x) \, \mathrm{d}x \right]. \tag{27}$$

Thus, the evolution of one mode is linked to that of every other mode. In the general case, this results in an infinite-dimensional system of equations, and is intractable. To implement it computationally, it is necessary to truncate the sum in eq. (26) after N terms. Then, the coupling between different modes is described by the N-dimensional matrix $\mathbf{M}$, defined such that

$$M_{nm} = \int_0^1 \rho(x) \sin(m\pi x) \sin(n\pi x) \, \mathrm{d}x. \tag{28}$$

In the case of a uniform string, this integral will vanish unless $n = m$, and so each mode behaves independently; however, if the density structure is not homogenous, energy exchange between modes can occur. This is exactly analogous to the manner in which material heterogeneities lead to mode coupling in the Earth (e.g. Woodhouse & Dahlen 1978). Defining $\mathbf{K}$ such that

$$K_{nm} = \frac{\delta_{nm}}{n^2 \pi^2} \tag{29}$$

where δ_{nm} represents a Kronecker delta, and representing the N time-varying coefficients, a_n, as the vector $\mathbf{a}$, we can write eq. (27) in the form

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \mathbf{a} \\ \dot{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} -\alpha \mathbf{K} & -2\mathbf{KM} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a} \\ \dot{\mathbf{a}} \end{pmatrix}. \tag{30a}$$

Here, $\mathbf{I}$ represents an identity matrix. Eq. (30a) is a system of coupled ordinary differential equations, and can readily be solved numerically using standard techniques, subject to the initial conditions

$$a_n(0) = 2 \int_0^1 y_0(x) \sin(n\pi x) \, \mathrm{d}x$$

$$\dot{a}_n(0) = 0. \tag{30b}$$

The resulting values of a_n can be used in conjunction with eq. (26) to model the behaviour of the string over time.

4.1.1 Model parametrization and partial derivatives

Given observations of the vibrations of a string, we wish to set up an inverse problem to determine its density structure. This requires us to select some method for parametrizing the density distribution, $\rho(x)$, typically by reference to some set of L spatial basis functions ($\psi_1(x), \psi_2(x), \ldots, \psi_L(x)$). This allows us to specify a continuous density model in terms of L model parameters, $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_L)$, such that

$$\tilde{\rho}(x) = \sum_{l=1}^{L} \rho_l \psi_l(x). \tag{31}$$

We use the notation $\tilde{\rho}(x)$ to denote a model for $\rho(x)$, in order to emphasize that the use of a finite set of basis functions imposes restrictions upon the range of density structures which may be represented. Using the theory set out in eqs (26)–(30b), it is straightforward to compute $\tilde{y}(x, t, \boldsymbol{\rho})$, the string vibrations corresponding to density model $\boldsymbol{\rho}$. Further, we can obtain partial derivatives with respect to individual model parameters, since

$$\frac{\partial \tilde{y}}{\partial \rho_\lambda} = \sum_{n=1}^{N} \frac{\partial \tilde{a}_n}{\partial \rho_\lambda} \sin(n\pi x), \tag{32}$$

and

$$\frac{\partial \tilde{a}_n}{\partial \rho_\lambda} = \frac{-1}{n^2 \pi^2} \left[\alpha \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \tilde{a}_n}{\partial \rho_\lambda} + 2 \sum_{m,l} \rho_l \frac{\mathrm{d}^2}{\mathrm{d}t^2} \frac{\partial \tilde{a}_m}{\partial \rho_\lambda} \right.$$
$$\times \int_0^1 \psi_l(x) \sin(m\pi x) \sin(n\pi x) \, \mathrm{d}x + 2 \sum_m \frac{\mathrm{d}^2 \tilde{a}_m}{\mathrm{d}t^2}$$
$$\left. \times \int_0^1 \psi_\lambda(x) \sin(m\pi x) \sin(n\pi x) \, \mathrm{d}x \right]. \tag{33}$$

Using $\boldsymbol{\Psi}^{(\lambda)}$ to denote the matrix with elements

$$\Psi_{mn}^{(\lambda)} = \int_0^1 \psi_\lambda(x) \sin(m\pi x) \sin(n\pi x) \, \mathrm{d}x \tag{34}$$

and noting that $\tilde{\mathbf{M}} = \sum_l \rho_l \boldsymbol{\Psi}^{(l)}$, we can then write down another coupled system of differential equations,

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \tilde{\mathbf{a}} \\ \dot{\tilde{\mathbf{a}}} \\ \partial \tilde{\mathbf{a}}/\partial \rho_\lambda \\ \partial \dot{\tilde{\mathbf{a}}}/\partial \rho_\lambda \end{pmatrix} = \begin{pmatrix} -\alpha \mathbf{K} & -2\mathbf{K}\tilde{\mathbf{M}} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -2\mathbf{K}\boldsymbol{\Psi}^{(\lambda)} & -\alpha \mathbf{K} & -2\mathbf{K}\tilde{\mathbf{M}} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} \tilde{\mathbf{a}} \\ \dot{\tilde{\mathbf{a}}} \\ \partial \tilde{\mathbf{a}}/\partial \rho_\lambda \\ \partial \dot{\tilde{\mathbf{a}}}/\partial \rho_\lambda \end{pmatrix}. \tag{35a}$$

Again, this can readily be solved, subject to the boundary conditions

$$\tilde{a}_n(0) = 2 \int_0^1 y_0(x) \sin(n\pi x) \, dx$$

$$\dot{\tilde{a}}_n(0) = 0$$

$$\left. \frac{\partial \tilde{a}_n}{\partial \rho_\lambda} \right|_{t=0} = 0$$

$$\left. \frac{\partial \dot{\tilde{a}}_n}{\partial \rho_\lambda} \right|_{t=0} = 0. \tag{35b}$$

Thus, for any set of model parameters $\boldsymbol{\rho}$, it is possible to compute synthetic data and the derivatives of those synthetics with respect to each model parameter. This makes it straightforward to set up a linearized inverse problem to determine the density structure of any string assumed to obey eq. (25).

4.2 Demonstration

We compute synthetic data for a string with density distribution

$$\rho(x) = 5 + 3x^2 + \cos(5\pi x) \tag{36}$$

and damping coefficient $\alpha = 1/2$. At $t = 0$, the string is at rest, with initial configuration

$$y_0(x) = e^{-8(5x-1)^2} \tag{37}$$

which represents a Gaussian wavepacket of width (*cf.* standard deviation) 0.05, centred on $x = 0.2$. We calculate the motion of the point at $x = 0.8$ over 60 time units, resulting in the waveform shown in Fig. 1. For these calculations, we truncate the summation in eq. (26) (and elsewhere) at $N = 100$; we have verified that higher terms are indeed negligible for present purposes. Note that the integrals in eqs (28) and (30b) can be solved numerically, and no model parametrization needs to be specified at this stage.

4.2.1 Recovery of density structure: least-squares inversion

Starting from this waveform, we wish to construct a model for the string's density structure. It is therefore necessary to decide upon a particular model parametrization. We use a one-dimensional grid, dividing the string into L equal-sized pieces, and assuming that each can be described by a constant density. Thus, our basis functions are defined

$$\psi_l(x) = \begin{cases} 1 & (l-1)/L < x < l/L \\ 0 & \text{otherwise} \end{cases}. \tag{38}$$
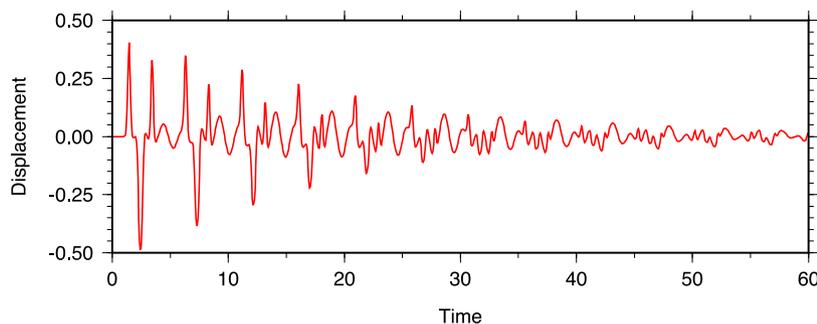
Note that this may not be the optimal choice: basis functions with global support are arguably more natural in conjunction with a normal-modes approach. However, for the purposes of this paper, the simplicity of a grid parametrization is desirable.

We choose $L = 20$, and assume that the overall mass of the string is known. We use this to construct an homogeneous initial model

$$\boldsymbol{\rho}_0 = (\bar{\rho}, \bar{\rho}, \dots, \bar{\rho}) \tag{39}$$

where $\bar{\rho}$ represents the mean value of the function $\rho(x)$ (eq. 36). As set out above, we can compute the synthetic data corresponding to this model, and the partial derivatives with respect to each model parameter. Thus, we can implement the least-squares algorithm, as set out in eq. (4b). We use Tikhonov regularization, $\mathbf{D} = \epsilon \mathbf{I}$, with $\epsilon = 0.1$. Note that it would be straightforward to include a mass constraint within the inversion (e.g. Menke 1989), although for the purposes of this illustration, we do not do so.

We perform five iterative updates to the model, tracking the misfit

$$m^2 = \frac{(\mathbf{d} - \mathbf{s})^{\mathrm{T}}(\mathbf{d} - \mathbf{s})}{\mathbf{d}^{\mathrm{T}}\mathbf{d}} \tag{40}$$

as the inversion proceeds. These are shown in Fig. 2, and demonstrate that the recovered model explains the available data almost perfectly. This model is also shown in Fig. 2, along with the 'true' density structure (eq. 36). We see that there is a good visual agreement between the two, although they do not match exactly: this is unsurprising, due to the limitations of a finite model parametrization. Nevertheless, it is evident that effective models of the string's density structure can be constructed using the data in Fig. 1.

4.3 An 'approximate' theory

The computation of the time-varying modal excitations $a_n(t)$ is complicated by the coupling between modes. However, inspection of the matrix $\mathbf{M}$ indicates that it is diagonally dominant: while density heterogeneities introduce energy exchange between modes, the strongest contribution comes from their effect on the individual modes. Neglecting off-diagonal terms allows each excitation coefficient to be determined independently (with corresponding derivatives, if required, by ignoring off-diagonal terms in $\boldsymbol{\Psi}^{(l)}$), greatly simplifying the computational procedure; an analytic solution to eq. (30) can readily be found. This is rather similar in conception to the 'self-coupling' approximation often adopted in global seismology, although there degeneracy effects are also present.

4.3.1 Symmetry considerations

In order to properly understand the effects of using this approximation within the inverse problem, it is necessary to briefly consider

Figure 1. The oscillations of a non-uniform string. Synthetic waveform representing motion of single point ($x = 0.8$) on a string of non-uniform density. A full description of the properties of the string is given in Section 4.2. All quantities are measured in arbitrary units.
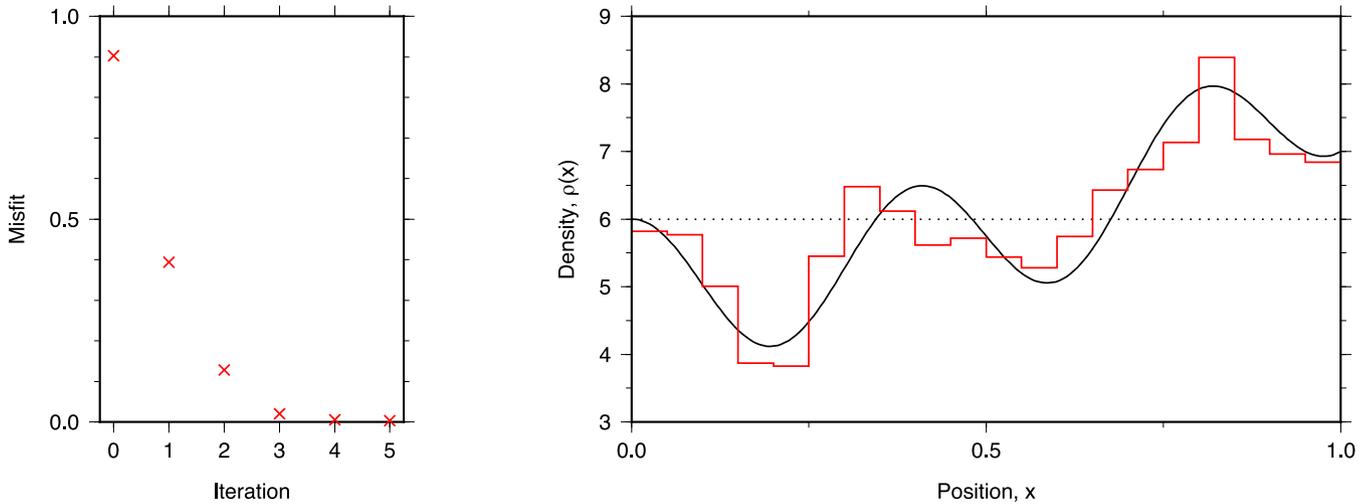
Figure 2. Least-squares inversion for string density structure. Left: iteration-by-iteration misfit (defined as in eq. 40) during least-squares inversion for the density variations along a vibrating string; after five iterations, almost perfect agreement is obtained between data and synthetics. Right: model recovered after five iterations (red), superimposed upon 'true' structure (black, solid). The initial model, at 'iteration 0', is the homogeneous density structure shown by the black dotted line. In this experiment, all quantities are computed in a theoretically complete manner, and no approximations are introduced except those associated with the use of a finite set of basis functions. Density and position are specified in arbitrary units.

the role of symmetry in the mode coupling equations. Any function $\rho(x)$ can be split into a purely symmetric part ρ_s, and a purely antisymmetric part, ρ_a, relative to a symmetry axis x_0

$$\rho(x) = \rho_s(x - x_0) + \rho_a(x - x_0) \tag{41}$$

where symmetry and antisymmetry have their usual definitions

$$\rho_s(x) = \rho_s(-x)$$
$$\rho_a(x) = -\rho_a(-x) \tag{42}$$

and taking x_0 as the midpoint of our string, i.e. $x_0 = 1/2$, eq. (28) can be expressed

$$M_{nm} = \int_0^1 \rho_s\left(x - \frac{1}{2}\right) \sin(n\pi x)\sin(m\pi x)\,\mathrm{d}x$$
$$+ \int_0^1 \rho_a\left(x - \frac{1}{2}\right) \sin(n\pi x)\sin(m\pi x)\,\mathrm{d}x \tag{43}$$

The function $\sin(n\pi x)$ may be either purely symmetric or purely antisymmetric about $x = 1/2$, depending on n, and in general the same applies to the product $\sin(n\pi x)\sin(m\pi x)$. However, in the case where $n = m$, both sinusoids necessarily share the same symmetry properties. Their product is always symmetric, and the second integrand in eq. (43) is always antisymmetric about $x = 1/2$. Thus, this second integral vanishes: M_{nn} depends only on the symmetric part of the density structure.

We therefore see that by making the approximation that off-diagonal terms in **M** and $\Psi^{(l)}$ are negligible, we lose all sensitivity to antisymmetric structure. A similar effect is seen when the self-coupling approximation is used for computing seismograms: the calculations only depend on the even-degree components in a spherical harmonic expansion of Earth structure. Conversely, where earth models are constructed based only on self-coupling approximations, odd-degree structure cannot be imaged; coupling between modes must be considered to access this (e.g. Resovsky & Ritzwoller 1995; Laske & Widmer-Schnidrig 2007).

4.3.2 Fully approximate inversion

This effect can readily be seen if we repeat the inversion of the data in Fig. 1, but instead use this 'self-coupling' approximation when computing both synthetics and partial derivatives. The results are shown in Fig. 3: a 70 per cent variance reduction is achieved over five iterations, implying that the recovered model explains aspects of the data, but it appears to bear little relationship to the 'true' density structure. It is also not an accurate depiction of the symmetric part of $\rho(x)$, as can be seen from Fig. 4, although there are some similarities between the two: in particular, the range of values spanned by the recovered model is similar to that of the symmetric part of the true model. Of course, this cannot be taken as a general rule. If this were a 'real' experiment, we suspect that any physical interpretation derived from these results would be substantially misled by the differences between recovered and true models.

4.3.3 Hybrid inversion

Fig. 3 also shows results for inversion of the same data using a 'hybrid' approach, where partial derivatives are computed using the 'self-coupling' approximation, but synthetics are calculated exactly. This leads to a very slight reduction in misfit compared to the fully approximate case, and the recovered model is almost unchanged. Although the theoretical framework used for computing synthetic data is more complete, the theoretically incomplete partial derivatives cannot make use of this information.

Since the starting model in Fig. 3 is symmetric, it might be thought that this is a special case: this may be an example where the difference between exact and approximate synthetics is limited. We therefore repeat the experiment using an initial model derived from the function $\rho_0 = 5 + 3x^2$, to match the gross behaviour of eq. (36). Again, we compare fully approximate and hybrid inversion schemes, with the results shown in Fig. 5. Unfortunately, the benefits of the hybrid approach remain limited, although the misfit values obtained using the hybrid scheme are significantly lower than those of the fully approximate case. This presumably arises because the exact synthetics used in hybrid inversion (and in computing
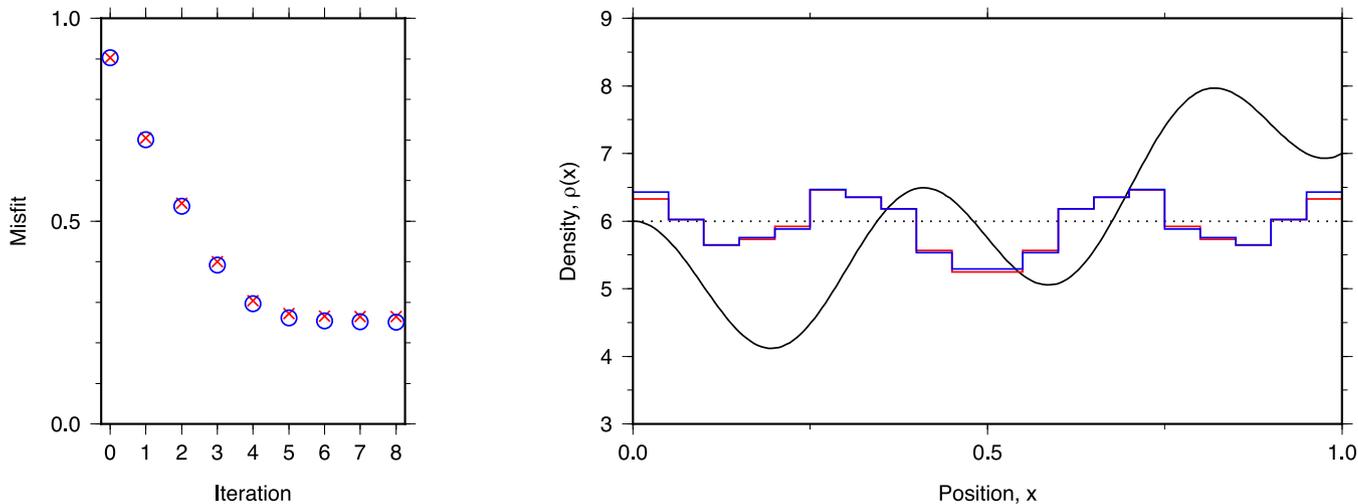
Figure 3. Comparison between fully approximate and hybrid inversion; compare Fig. 2. Left: iteration-by-iteration misfit for fully approximate (red crosses) and hybrid (blue circles) inversion of the data given in Fig. 1, adopting the 'self-coupling' approximation described in Section 4.3. Right: recovered models after eight iterations of the fully approximate (red) and hybrid (blue) inversion schemes. The initial model is given by the black dotted line; the black solid line represents the 'true' model used when generating the data. It is evident that the use of a more complete synthetic calculation does not significantly alter the recovery.
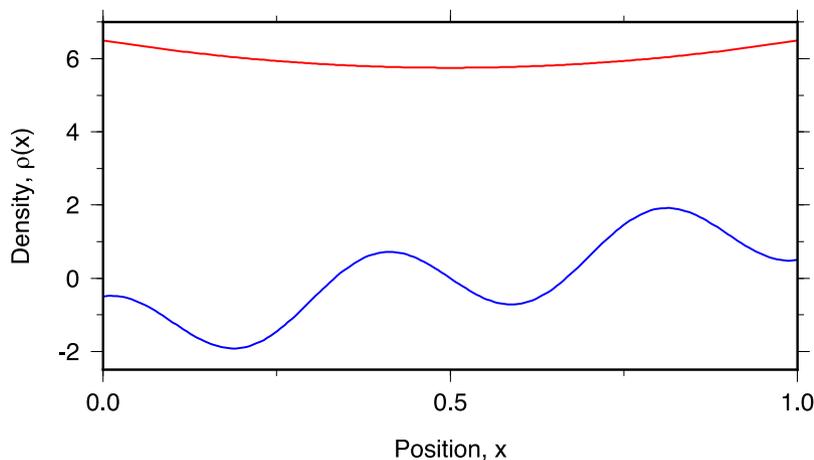
Figure 4. Symmetric (red) and antisymmetric (blue) parts of $\rho(x)$ (eq. 36). The symmetry axis is taken to be $x = 1/2$.

the corresponding misfit) incorporate the antisymmetric part of the density structure, which is lost to the approximate synthetics. However, the approximate derivatives remain largely unable to access this information.

4.4 An alternative approximation: neglecting attenuation

To explore whether these effects are specific to the case of 'self-coupling', we consider making a different approximation, whereby we neglect the effects of attenuation. Setting $\alpha = 0$, we solve eq. (35) allowing coupling between all modes, and attempt to determine the string's density structure from the data in Fig. 1. Again, we use the term 'fully approximate' inversion to imply that attenuation is neglected for both synthetics and partial derivatives; hybrid inversion uses the correct value of α in computing synthetics, but $\alpha = 0$ for partial derivatives. Results are shown in Fig. 6.

In this case, the results of hybrid inversion can probably be regarded as an improvement over those obtained in the fully approximate setting. This assessment is not wholly objective: we do not attempt to quantify the quality of recovered models, which is likely

to be application-specific. However, inspection of Fig. 6 shows that the fully approximate algorithm significantly overestimates the amplitude of the minimum and maximum density anomalies. The effect is still present in the hybrid case, but is much less pronounced. Elsewhere, neither model perfectly represents the true structure, and it is not obvious that one is consistently better than the other. Nevertheless, the misfit values obtained during hybrid inversion are much improved over those when approximate synthetics are used.

4.4.1 Geometrical considerations

In order to illustrate one perspective on why a hybrid approach brings greater benefits in conjunction with the 'non-attenuating' approximation than the 'self-coupling' approximation, we turn to consider geometrical factors. In Section 3.1.1, we argued that any change to the forward calculation can only impact inversion results to the extent that it lies within the image space of the inverse operator. For the case of hybrid inversion, Section 3.2.3—and specifically, eq. (17)—indicates that any benefits must be driven by the quantity $\boldsymbol{\Delta}_{i+1}^{\text{fwd}} = \mathcal{S}_{\mathbf{h_i}}^{-g}(\mathbf{s}(\mathbf{h_i}) - \mathbf{t}(\mathbf{h_i}))$, where $\mathcal{S}_{\mathbf{h_i}}^{-g}$ represents the approximate
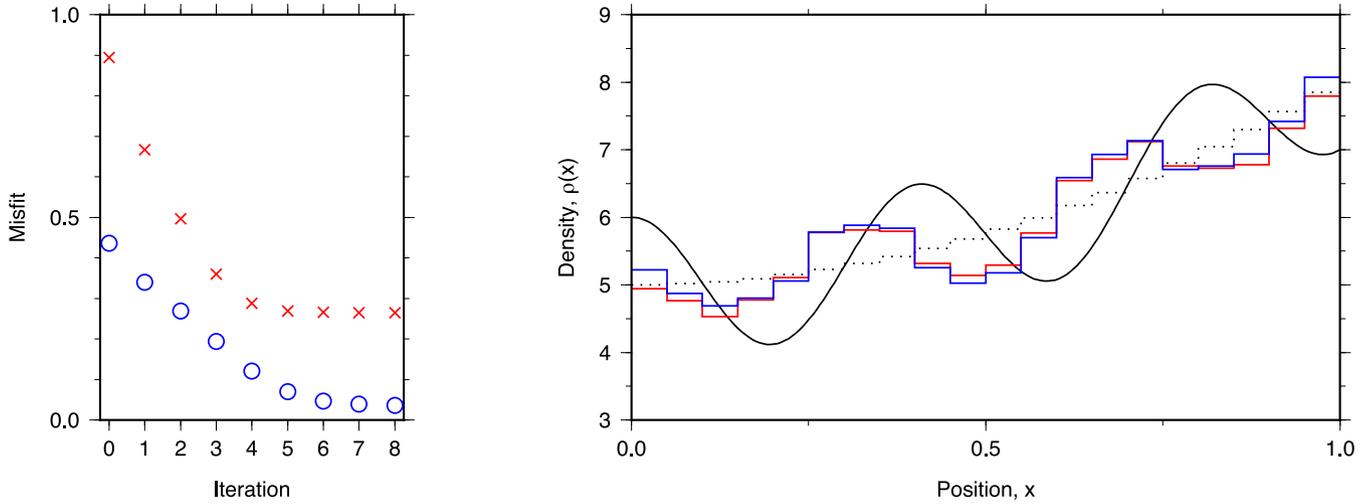
Figure 5. Comparison between fully approximate and hybrid inversions, non-uniform initial model. See also Figs 2 and 3. Left: misfit evolution for fully approximate (red crosses) and hybrid (blue circles) inversion of the data in Fig. 1. Right: recovered models for fully approximate (red) and hybrid (blue) schemes, after eight iterations; again, the 'self-coupling' approximation was used. Inversion began with a non-uniform model (black, dotted); the true model is represented by the black solid line. The model obtained via hybrid inversion appears to give a much lower misfit than that in the fully approximate case, yet there is little difference between results.
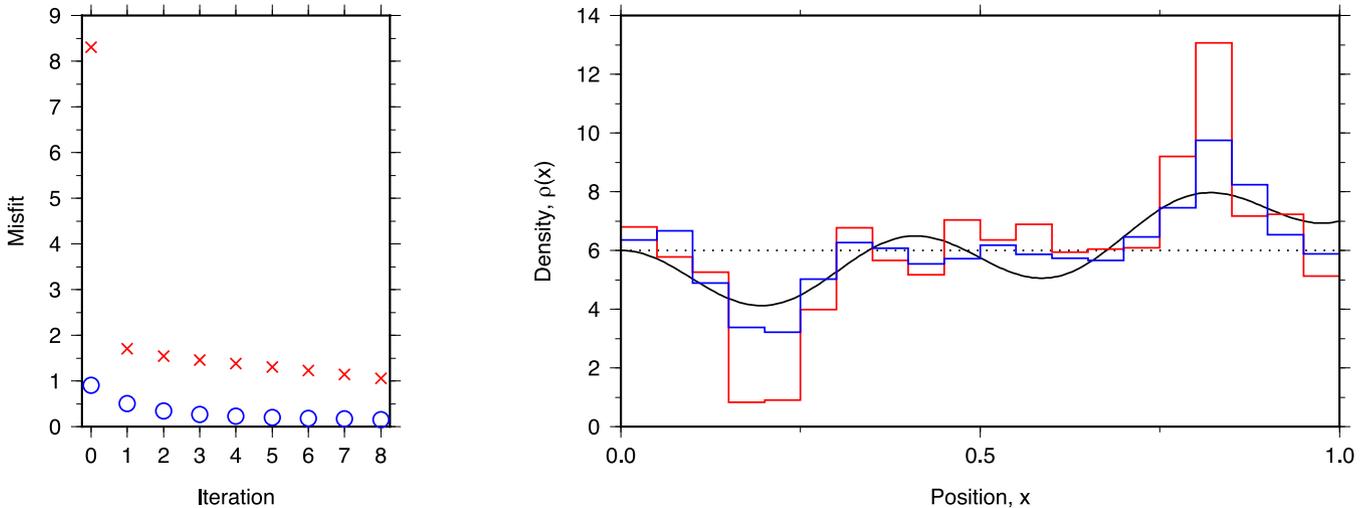
Figure 6. Comparison between hybrid and approximate inversions, attenuation neglected. Compare especially Fig. 3; note that vertical axes are considerably extended relative to previous plots. Left: misfit evolution for fully approximate (red crosses) and hybrid (blue circles) inversion, where the 'approximation' involves ignoring the effects of attenuation (i.e. assuming $\alpha = 0$, but otherwise solving eq. (30) exactly). Right: recovered models after eight iterations of fully approximate (red) and hybrid (blue) inversion. Also shown are the true model (black, solid) and the initial model (iteration 0; black, dotted).
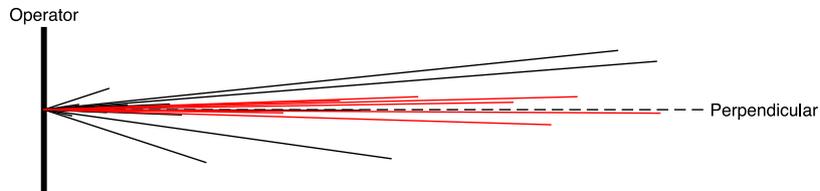
Figure 7. Geometrical relationships between approximate inverse operators and omitted forward physics. Sketch of relationship between the data-space singular vectors of the (approximate) inverse operator ('Operator'), and the difference between exact and approximate synthetics. Lines are constructed by computing the angle between this difference and the ith singular vector (θ_i, eq. 44), and scaling line length according to the corresponding singular value. Red denotes 'self-coupling' approximation (Section 4.3), while black denotes the 'non-attenuating' case (Section 4.4). See Section 4.4.1 for more detailed explanation.

inverse operator, $s(\mathbf{h_i})$ is the approximate synthetics and $t(\mathbf{h_i})$ the exact synthetics. If the 'improvement' associated with exact synthetics, $(s(\mathbf{h_i}) - t(\mathbf{h_i}))$, does not lie in the subspace accessible with the imaging operator, $\mathcal{S}_{\mathbf{h_i}}^{-g}$, it cannot influence inversion results. Fig. 7 illustrates that this situation arises to a greater extent for

the 'self-coupling' approximation than for the 'non-attenuating' approximation.

We use $\mathbf{A_s}$ to denote the matrix of partial derivatives evaluated for model $\mathbf{h_i}$ using the 'self-coupling' approximation and $\mathbf{A_n}$ for the 'non-attenuating' approximation. To simplify comparisons, we

choose $\mathbf{h}_i$ to be the uniform initial model, $\boldsymbol{\rho}_0$. The inverse operators, $\mathcal{S}_{\mathbf{h}_i}^{-g}$, are constructed as in eq. (4), with regularization $\mathbf{D} = 0.1\mathbf{I}$ (as already described). Each inverse operator can be expressed using an SVD, as in eq. (5). In the present case, we have $L = 20$ model parameters, and thus at most L non-zero singular values. For the 'self-coupling' case we denote these by $\{\sigma_1^{(s)}, \sigma_2^{(s)}, \ldots, \sigma_L^{(s)}\}$, with corresponding right-singular vectors $\{\mathbf{v}_1^{(s)}, \mathbf{v}_2^{(s)}, \ldots, \mathbf{v}_L^{(s)}\}$. Note that since we are constructing the SVD of the inverse operator, these vectors lie in the data space. For the operator in the 'non-attenuating' case, the equivalent (but non-identical) quantities are denoted $\{\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots, \sigma_L^{(n)}\}$ and $\{\mathbf{v}_1^{(n)}, \mathbf{v}_2^{(n)}, \ldots, \mathbf{v}_L^{(n)}\}$. For notational brevity, we introduce $\mathbf{e}^{(s)} = \mathbf{s}(\mathbf{h}_i) - \mathbf{t}(\mathbf{h}_i)$ to represent the difference between exact and approximate synthetics when $\mathbf{s}$ is calculated using the 'self-coupling' approximation and $\mathbf{e}^{(n)}$ for the equivalent quantity in the 'non-attenuating' case.

With these quantities defined, it is straightforward to compute the angle between any singular vector of the inverse operator and the forward-modeling error eliminated by switching from approximate to exact calculations,

$$\theta_i^{(s,n)} = \arccos\left(\frac{\mathbf{v}_i^{(s,n)} \cdot \mathbf{e}^{(s,n)}}{\sqrt{\mathbf{e}^{(s,n)} \cdot \mathbf{e}^{(s,n)}}}\right), \qquad (44)$$

where we have noted that the singular vectors are, by definition, normalized. The closer these angles are to a right angle, the less the influence on inversion results. Fig. 7 sketches these quantities in the 2-D plane for the two cases, with the 'self-coupling' approximation in red; the length of each line is determined by the corresponding singular value. It is evident that in the 'self-coupling' case, $\mathbf{e}^{(s)}$ is almost orthogonal to the hyperplane defined by the imaging operator, so that there is little difference in result between fully approximate and hybrid inversions. When the 'non-attenuating' approximation is used, $\mathbf{e}^{(n)}$ is somewhat less orthogonal to the operator, and there is greater benefit associated with a change to using exact synthetics. Thus, our experimental observations of relative algorithmic performance are in accordance with the theoretical considerations described in Section 3.

5 CONCLUDING REMARKS

In this paper, we have set out a fairly general statement of linearized inverse theory, and described how differences between two iterative inversion algorithms manifest themselves in the recovered images. This can be used to assess how particular choices or assumptions—such as on model parametrization, data processing, or the theoretical basis for performing calculations—affect results. This may be used to understand why two images that purport to describe the same system are not identical, permitting a more principled approach to interpretation. Alternatively, one can compare a specific imaging framework to an idealized algorithm in which no approximations or assumptions need be made. Again, this may assist interpretation, by providing insight into any distortions or limitations in recovered models. This knowledge, in turn, can enable the optimization of the imaging process itself.

Our analysis identifies three factors that contribute to image differences (as in eq. 9): differences between the forward and inverse operators used; the fact that two iterative algorithms may follow distinct paths through model space; and effects arising from the (typically) imperfect resolution of the imaging operators. If one chooses to compare two specific imaging algorithms, all operators will be known, and the relative contributions of these terms can be quantitatively assessed. However, any insights may not apply be-

yond the particular comparison performed. In the present work, we have adopted a more abstract approach, and explored some general properties of linearized inversion.

Taking such a general approach, the relative performance of two imaging algorithms can be understood in terms of the different null and image spaces of the various operators involved. In particular, it is important to recognize that even 'small' changes to these operators can nevertheless generate substantial changes to these subspaces—especially where they are defined in high-dimensional spaces and where the null space is large compared to the image space. Both of these conditions are the norm in many geophysical inverse problems. This can lead to counter-intuitive effects: an approximation that seems 'small' in magnitude may nevertheless substantially alter the geometric relationship between operators, and therefore generate a larger effect on the recovered image than might otherwise be supposed.

A case in point is the concept of 'hybrid' inversion, where an accurate but expensive forward calculation is used in combination with an inverse operator obtained more cheaply, using approximations. It is typically assumed that the accuracy of the recovered image is primarily influenced by the accuracy of the forward code, and thus that hybrid inversion will yield a high-quality model. However, it is clear that this is not—as a general statement—true. We see no reason to suppose that approximations in the inverse operator are necessarily less important than those in the forward operator, unless this can be justified in a specific case by reference to the particular approximations envisaged. It is also noteworthy that physically motivated approximations will often have strong geometrical effects upon the imaging operators, substantially altering their null and image spaces. It is therefore quite possible for many of the 'benefits' of the accurate forward model to lie in the null space of the approximate imaging operator, and to be lost. Such behaviour is clearly seen in the example presented in Section 4, where use of a hybrid approach does not lead to a markedly better model, despite significant improvements in data fit. We emphasize that we are not arguing that hybrid approaches should be entirely discredited: in some circumstances they may lead to higher quality models. However, such a claim must be substantiated through analysis of the specific approximations that are being made (or avoided).

In recent years, geophysicists have benefited immensely from advances in data quality and quantity, and in the theoretical completeness of available computational codes. This has enabled the production of models with increasingly high apparent resolution, yet two representations of the same system rarely agree at a detailed level. Some portion of these differences represent the inherent uncertainties associated with attempting to represent a physical system from noisy, incomplete measurements. The remainder arise from subtle differences in how the 'imaging problem' has been posed in each case. Understanding and quantifying these two effects is vital if models are to be interpreted properly. Indeed, one might reasonably question the value in producing ever-more-detailed models without first reconciling or explaining such discrepancies; all experience suggests that the problem will only become more pronounced as detail increases. The framework presented here may offer one route toward tackling these questions, although it is clear that much remains to be done.

This paper has focused exclusively on linearized inverse theory, whereby imaging algorithms are framed in terms of linear algebra, and where each iterative improvement in the model is constructed by assuming a linear relationship between observables and parameters until a single optimal solution is found. This may not be appropriate in all cases: non-linearities in the true forward problem may result

in poor performance, and where large null spaces exist, the value of a single image may be limited. In such circumstances, other approaches to imaging—typically, those based on testing models at random, to characterize the range of images compatible with data—are required. Such techniques pose their own challenges, and choices, assumptions and approximations continue to occur. However, in some ways, the 'imaging operator' for the non-linear case is much simpler and readily understood than that for iterative, linearized inversion. As computational resources grow and non-linear approaches become increasingly feasible, the interpretational challenges associated with linearized inversion techniques may weigh increasingly heavily in the balance.

ACKNOWLEDGEMENTS

REFERENCES

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178,** 1411–1436.

Boschi, L. & Dziewoński, A., 1999. High- and low-resolution images of the Earth's mantle: implications of different approaches to tomographic modeling, *J. geophys. Res.,* **104,** 25 567–25 594.

Bozdağ, E. & Trampert, J., 2008. On crustal corrections in surface wave tomography, *Geophys. J. Int.,* **172,** 1066–1082.

Cormier, V., 2007. Forward modeling/synthetic body wave seismograms, in *Seismology and Structure of the Earth,* vol. 1 of Treatise on Geophysics, pp. 157–189, eds Romanowicz, B. & Dziewoński, A., Elsevier.

de Wit, R.W.L., Trampert, J. & van der Hilst, R.D., 2012. Toward quantifying uncertainty in travel time tomography using the null-space shuttle, *J. geophys. Res.,* **117**(B3), doi:10.1029/2011JB008754.

de Wit, R., Käufl, P., Valentine, A. & Trampert, J., 2014. Bayesian inversion of free oscillations for Earth's radial (an)elastic structure, *Phys. Earth planet. Inter.,* **237,** 1–17.

Dosso, S., Holland, C. & Sambridge, M., 2012. Parallel tempering for strongly nonlinear geoacoustic inversion, *J. acoust. Soc. Am.,* **132,** 3030–3040.

Dziewonski, A., Chou, T.-A. & Woodhouse, J., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. geophys. Res.,* **86,** 2825–2852.

Fichtner, A. & Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.,* **185,** 775–798.

Fichtner, A. & Trampert, J., 2011b. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **187,** 1604–1624.

French, S., Lekić, V. & Romanowicz, B., 2013. Waveform tomography reveals channeled flow at the base of the oceanic asthenosphere, *Science,* **342,** 227–230.

Golub, G. & van Loan, C., 2012. *Matrix Computations,* 4th edn, John Hopkins University Press.

Grcar, J., 2010. Spectral condition numbers of orthogonal projections and full rank linear least squares residuals, *SIAM J. Matrix Anal. Appl.,* **31,** 2934–2949.

Käufl, P., Valentine, A., O'Toole, T. & Trampert, J., 2014. A framework for fast probabilistic centroid–moment-tensor determination—inversion of regional static displacement measurements, *Geophys. J. Int.,* **196,** 1676–1693.

Komatitsch, D., Ritsema, J. & Tromp, J., 2002. The spectral-element method, Beowulf computing, and global seismology, *Science,* **298,** 1737–1742.

Laske, G. & Widmer-Schnidrig, R., 2007. Normal mode and surface wave observations, in *Treatise on Geophysics,* pp. 67–125, Elsevier.

Lekić, V. & Romanowicz, B., 2011. Inferring mantle structure by full waveform tomography using the spectral element method, *Geophys. J. Int.,* **185,** 799–831.

Lévêque, 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.

Meju, M. & Hutton, V., 1992. Iterative most-squares inversion: application to magnetotelluric data, *Geophys. J. Int.,* **108,** 758–766.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press, New York.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L. & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion with multicomponent data: from theory to practice, *Leading Edge,* **32,** 1040–1054.

Pratt, R., Shin, C. & Hicks, G., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.,* **133,** 341–362.

Resovsky, J. & Ritzwoller, M., 1995. Constraining odd-degree Earth structure with coupled free oscillations, *Geophys. Res. Lett.,* **22,** 2301–2304.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.,* **138,** 479–494.

Schaeffer, A. & Lebedev, S., 2013. Global shear speed structure of the upper mantle and transition zone, *Geophys. J. Int.,* **194,** 417–449.

Schaeffer, A. & Lebedev, S., 2015. Global heterogeneity of the lithosphere and underlying mantle: a seismological appraisal based on multimode surface-wave dispersion analysis, shear-velocity tomography, and tectonic regionalization, in *The Earth's Heterogeneous Mantle,* pp. 3–46, eds Khan, A. & Deschamps, F., Springer.

Stewart, G., 1977. On the perturbation of pseudo-inverses, projections and linear least squares problems, *SIAM Rev.,* **19,** 634–662.

Tarantola, A. & Valette, B., 1982. Inverse problems = Quest for information, *J. Geophys.,* **50,** 159–170.

Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: possible artifacts in seismic tomography, *Science,* **271,** 1257–1260.

Trampert, J., Fichtner, A. & Ritsema, J., 2013. Resolution tests revisited: the power of random numbers, *Geophys. J. Int.,* **192,** 676–680.

Valentine, A. & Trampert, J., 2012. Assessing the uncertainties on seismic source parameters: towards realistic error estimates for centroid–moment-tensor determinations, *Phys. Earth planet. Inter.,* **210–211,** 36–49.

Valentine, A. & Woodhouse, J., 2010. Reducing errors in seismic tomography: combined inversion for sources and structure, *Geophys. J. Int.,* **180,** 847–857.

van der Sluis, A., 1975. Stability of the solutions of linear least squares problems, *Numer. Math.,* **23,** 241–254.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics,* **74,** WCC127–WCC152.

Woodhouse, J. & Dahlen, F., 1978. The effect of a general aspherical perturbation on the free oscillations of the Earth, *Geophys. J. R. astr. Soc.,* **53,** 335–354.