ORIGINAL PAPER

Multi-physics adjoint modeling of Earth structure: combining gravimetric, seismic, and geodynamic inversions

Georg S. Reuber[1,2,3] · **Frederik J. Simons**[4,5]

© The Author(s) 2020

Abstract

We discuss the resolving power of three geophysical imaging and inversion techniques, and their combination, for the reconstruction of material parameters in the Earth's subsurface. The governing equations are those of Newton and Poisson for gravitational problems, the acoustic wave equation under Hookean elasticity for seismology, and the geodynamics equations of Stokes for incompressible steady-state flow in the mantle. The observables are the gravitational potential, the seismic displacement, and the surface velocity, all measured at the surface. The inversion parameters of interest are the mass density, the acoustic wave speed, and the viscosity. These systems of partial differential equations and their adjoints were implemented in a single Python code using the finite-element library FeNICS. To investigate the shape of the cost functions, we present a grid search in the parameter space for three end-member geological settings: a falling block, a subduction zone, and a mantle plume. The performance of a gradient-based inversion for each single observable separately, and in combination, is presented. We furthermore investigate the performance of a shape-optimizing inverse method, when the material is known, and an inversion that inverts for the material parameters of an anomaly with known shape.

✉ Georg S. Reuber
    reuber@uni-mainz.de

1   Institute of Geosciences, Johannes Gutenberg-University, Mainz 55128, Germany

2   Max-Planck Graduate Center, Mainz, Germany

3   Mainz Institute of Multiscale Modelling (M3ODEL), Johannes Gutenberg-University, 55128 Mainz, Germany

4   Department of Geosciences, Princeton University, Princeton, NJ 08544, USA

5   Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA

🖄 Springer

1 Introduction

Geophysics, both a systematic framework and a method for Earth exploration (Brown and Slawinski 2017) is traditionally divided into subdisciplines around distinct research methods, each focusing on a different observable, and making inferential statements about Earth properties through separate inversions for subsurface structure. In this paper we discuss the possibilities and advantages of conducting inversions that combine multiple physical observables in one single mathematical framework.

1.1 Forward and inverse problems in global geophysics

All geophysics, like ancient Gaul, is divided into three parts, one of which is the domain of potential-field methods (gravity, geomagnetism, geo-electricity), the other that of seismic wave phenomena, geodynamics occupying the third. All these differ from each other in formalism, conventions, and best practices. Seismology is the most direct of all of these approaches, because their sources—earthquakes—reach furthest down inside the Earth. The seismic wavefield thus rather directly samples its interior structure, carrying information back up to the receiver. In contrast, gravitational or geodynamic observations are perennially confined to the two-dimensional bounding surface of the volume whose structure and properties we seek to illuminate.

Inverse modelling in geophysics is classically divided into two distinct but overlapping realms. One of them aims to image the subsurface, focusing on material contrasts and high-frequency— "sharp"—structure. The second purports to constrain the parameters of the imaged subsurface structure, resolving material properties and low-frequency—"smooth"—structure. In seismology, where the distinction is most relevant (Mora 1988, 1989; Woodward 1992), one speaks of "migration" versus "tomography"—"inversion", by any other name (Nolet 2008; Schuster 2017).

Geophysical observables that are measured at Earth's surface ultimately derive from partial differential equations (PDEs), which lend themselves to theoretical analysis (Mead and Ford 2020; Michel and Simons 2017) and numerical simulation. One can synthetically create measurements based on an initial model—a starting guess for the subsurface structure—typically, but not exclusively, as we shall see, a "low-wavenumber" smooth model upon which "high-wavenumber" sharp contrasts are sought to be superposed (Bunks et al. 1995; Yuan and Simons 2014).

1.2 Geophysical observables and governing equations

In this work we consider three geophysical observables and their governing equations: (i) the gravitational acceleration and potential derived from Poisson's equation, (ii) the acoustic wave field controlled by the linear wave equation, and (iii) the mantle flow field governed by the steady-state non-inertial Stokes equations for mass and momentum conservation in incompressible media together with Poisson's equation.

The *gravitational acceleration*, the gradient of Newton's potential, is sensitive to mass density via Poisson's equation (Kellogg 1967). Inverting for Earth's density structure to match observed gravity anomalies, disturbances, potential anomalies, or

equipotential undulations (Blakely 1995; Hofmann-Wellenhof and Moritz 2006) is an ill-posed problem (Xu 1992a, b). Without prior information it is generally impossible to uniquely resolve the location and strength of the causative anomalies (Chao 2005; Dorman and Lewis 1970; Oldenburg 1974; Parker 1973), and the null-space with "anharmonic" density distributions remains inaccessible without non-gravimetric constraints (Fischer and Michel 2012; Michel 2005; Michel and Fokas 2008; Michel and Orzlowski 2015).

The second measurable is the *surface displacement* caused by the propagation of elastic or acoustic waves, e.g., after an earthquake (Aki and Richards 1980; Dahlen and Tromp 1998) or active probing (Sheriff and Geldart 1995; Yilmaz 2001). For simplicity we only look at the *pressure* caused by acoustic waves, inverting for the wave speed, which combines information on Hookean elastic moduli and mass density. The inversion of the acoustic wave equation cannot precisely discriminate the cause for the observed differences in the wave speed, but it is able to relatively accurately image the subsurface location of the anomalies (Gauthier et al. 1986; Tarantola 1984a, b).

The third measurable is the *surface velocity* caused by the motion of highly viscous solid rock in the subsurface (Glatzmaier 2014; Schubert et al. 2001), measured on geologic time scales or contemporaneously, e.g., by global navigation satellite systems (GNSS) such as the Global Positioning System (GPS). The material flow in the Earth's interior can be modeled by the Stokes equations (Gerya 2019; Kennett and Bunge 2008), which are non-linear. The inversion of surface velocities is characterized by a rather low sensitivity to the location of the driving forces, but it can with some accuracy infer the density and viscosity of subsurface anomalies, especially combined with inference from seismology (Conrad et al. 2013; Forte and Mitrovica 2001; Harig et al. 2010; Lithgow-Bertelloni and Richards 1998).

Newton, *Hooke*, and *Stokes*: by these mnemonics we designate the three different integro-differential forward operators that map different portions of parameter space onto distinct types of observations. The individual inverse problems have been addressed by many authors. Their combination, and, specifically, the relative value of combining observations, is the subject of this paper. Within the broader context of the numerical solution of inverse problems in geophysics, we unify approaches for these three very different problems, and discuss how to combine and contrast them, highlighting the trade-offs between their various sensitivities.

1.3 Geological test scenarios

We consider three simplified geological scenarios, (i) a sinking or "falling" block (Morgan 1965), which can be interpreted as a Rayleigh-Taylor instability (Conrad and Molnar 1997), e.g., representing a lithospheric drip or a rising magma chamber, (ii) a "subduction" setting (Stern 2002), described by a denser, higher-viscosity and colder, and, by inference, higher wave-speed "slab" embedded in a background mantle, and (iii) a buoyant, bell-shaped anomaly at the bottom of the domain embedded in a background mantle, such as can be interpreted as a simplified version of a mantle "plume" (Condie 2001; Morgan 1971).

1.4 Guide to the paper

We study the inverse problem by (i) mapping out the least-squares cost function for each of the attendant observables separately, and combined, to analyse how a combination of observations results in a more unique or steeper cost function shape. This is done for two material parameters that are defined on geometrically fixed subsets. We furthermore calculate (ii) the pointwise gradient fields of the forward solutions with respect to the solution variables, and (iii) the gradients of a cost function with respect to the material parameters for the inversion.

We limit our scope to gradient-based inversions, including possible regularization, discussing their use in resolving the geometry and amplitude of subsurface anomalies. We do not consider operator inversions (Michel and Simons 2017; Mead and Ford 2020), subspace methods (Geng et al. 2020) or statistical frameworks (Fichtner and Simutė 2018) that invert for anomalous subsurface structure by exploring the cost-function space by a combination of statistical and deterministic methods. Deterministic gradient-based methods such as those presented here can only find the minimum of a function that is closest to the initial guess, which is not necessarily the *global* minimum. However, they enjoy the advantage of a relatively low computational numerical cost.

We present the forward and adjoint equations of each of the PDEs used to simulate the three distinct geophysical observables. For the analysis we require the gradients of the material parameters with respect to the solutions. To this end we derive the adjoint equations for all three PDEs. The adjoint gradient calculation is independent of the number of parameters, only requiring one additional linear solution, thus making it possible to perform field-based inversions. We develop the adjoint system of equations using the formal Lagrangian approach following Tröltzsch (2010). This approach results in the weak forms of the forward, adjoint, and gradient equations, which are discretized using FEniCS (Alnæs et al. 2015; Logg et al. 2012), a high-level Python or C++ framework to efficiently develop finite-element (FE) codes using PETSc (Balay et al. 2019), a management suite of highly optimized parallel data structures which gives access to direct, as well as iterative solvers.

Lastly, we suggest a simple geometrical shape optimization algorithm that does not require shape discretization as boundary term. Such an algorithm can be useful if sufficient laboratory constraints on, e.g., rock type exist, such that the material parameters are well known and the shape of the anomaly is of dominant interest.

2 Theoretical framework

All of our considerations will be explicitized for a two-dimensional bounded domain of Cartesian space, $\Omega \in \mathbb{R}^2$, with boundary $\delta\Omega$, and space coordinates $\mathbf{r}$.

2.1 Gravity

Newton's gravitational potential, corrected for first-order effects such as the Earth's equilibrium shape and its rotation, varies continuously across the Earth's surface,

reflecting inhomogeneities in the mass density of the interior of the Earth (Blakely 1995). Additionally accounting for an independently known or presumed regional background mass distribution, we consider the relation (Hofmann-Wellenhof and Moritz 2006) between a localized density 'perturbation', $\rho(\mathbf{r})$, i.e. a mass anomaly at depth within the domain Ω, and the 'anomalous' potential, $u(\mathbf{r})$, that it generates, and to which we have access at the measurement surface, here presumed to be the flat upper boundary $\Omega_{\text{obs}} = \delta\Omega_{\text{top}}$:

$$\nabla^2 u = 4\pi G \rho \quad \text{within } \Omega, \tag{1}$$
$$u = u_{\text{obs}} \quad \text{on } \delta\Omega_{\text{top}}. \tag{2}$$

The unknown mass density ρ is in $\text{kg}\,\text{m}^{-3}$, the measurable gravitational potential u is in $\text{m}^2\,\text{s}^{-2}$, the universal gravitational constant $G = 6.674 \times 10^{-11}\ \text{m}^3\,\text{kg}^{-1}\,\text{s}^{-2}$, and ∇^2 is Laplace's operator. Obtaining the density distribution $\rho(\mathbf{r})$ from boundary observations of the harmonic potential $u(\rho)$ through Poisson's relation, Eq. (1), is the well-studied non-unique, exponentially ill-posed problem of 'inverse gravimetry' (Freeden and Nashed 2018; Michel and Fokas 2008).

We seek to minimize the quadratic misfit functional, subscripted N for "Newton",

$$F_N(u) = \frac{1}{2} \int_{\Omega_{\text{obs}}} \left[u_{\text{obs}} - u(\rho) \right]^2 dx, \tag{3}$$

an integral over the linear dimension x that defines the top boundary, $\delta\Omega_{\text{top}}$. In practice we will restrict our observations to discrete points on the upper surface $\delta\Omega_{\text{top}}$, representing surficial sensor locations.

We now formulate the adjoint version of this problem. We minimize the functional in Eq. (3) subject to the elliptic PDE constraint in Eq. (1). Following, e.g., Martin et al. (2012), Plessix (2006), Tröltzsch (2010), or Virieux and Operto (2009), we calculate the first, second and third variation of the Lagrangian,

$$\mathscr{L}_N = \frac{1}{2} \int_{\Omega_{\text{obs}}} \left[u_{\text{obs}} - u \right]^2 dx - \int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\Omega} 4\pi G \rho \, v \, dx, \tag{4}$$

which we recognize as the sum of $F_N(u)$, the function that is to be minimized, and the weak form of the forward problem that acts as a constraint to the minimization, which is obtained (Zienkiewicz 1977) by integrating Eq. (1) over Ω after multiplication by an appropriate test function v and its gradient ∇v.

The first variation of Eq. (4), with respect to the test function, returns the weak form of the forward problem in Eq. (1).

The second variation, with respect to the solution u, yields the 'adjoint equation', to be solved for the adjoint field v using a new set of test functions $\hat{u}$,

$$\frac{\partial \mathscr{L}_N}{\partial u}(\hat{u}) = \int_{\Omega_{\text{obs}}} \left[u_{\text{obs}} - u \right] \hat{u} \, dx - \int_{\Omega} \nabla v \cdot \nabla \hat{u} \, dx = 0. \tag{5}$$

Since u is the solution of the forward problem, it is clear that the forward equation must be solved before the adjoint equation, which, however, remains strictly linear. A useful interpretation of the adjoint equation is that it solves the original PDE but with the mismatch between observations and forward predictions as a source term.

The third variation is with respect to the design variable, here: the density ρ. This results in the weak form of the gradient of the misfit function,

$$\frac{\partial \mathscr{L}_N}{\partial \rho}(\hat{\rho}) = -\int_\Omega 4\pi G \, v \, \hat{\rho} \, dx = \frac{dF_N}{d\rho}. \tag{6}$$

Here, v is the solution of the adjoint equation, Eq. (5).

After discretization within Ω, the gradient can be used by the inversion procedure, e.g., via gradient descent, whereby the density is updated iteratively until convergence, according to the line search

$$\rho_{i+1} = \rho_i - \alpha_i \frac{dF_N}{d\rho}, \tag{7}$$

with i the iteration counter. We choose α_i to evolve according to:

$$\alpha_i = \zeta \frac{a_0}{\max\left(|dF_N/d\rho|_{i=0}\right)}, \tag{8}$$

where a_0 is the initial perturbation of the parameter, and ζ is an increment that is increased in case the gradient update was successful in terms of decreasing the misfit function, or decreased, if not.

More advanced techniques such as quasi-Newton methods improve convergence by supplying the Hessian of the problem (Fichtner and Trampert 2011; Ma and Hale 2012), the curvature of $F_N(u)$. For the simple test cases presented here, we save ourselves the computational burden.

The non-uniqueness of the gravity inversion problem is well-known (Oldenburg 1974). The spatial extent, depth of burial, and density contrast of the perturbing body trade off in producing identical disturbing potentials at the surface. While we treated the gravity inversion problem for subsurface structure imaging first, we envisage it as the second step to constrain the density of an anomaly whose approximate location can be independently imaged, e.g., by the seismic method discussed in the next section.

2.2 Seismology

Seismology is undoubtedly the major imaging tool in geophysics, both for exploration (Claerbout 1992) and whole-Earth structure (Nolet 2008). Whether excited by artificial or natural sources, diffusive forcings or earthquakes, the linearized balance between infinitesimal perturbations of stress and strain, within the constitutive framework of Hookean elasticity, results in the propagation of waves whose speeds are controlled by the distribution of mass density and stiffness parameters inside the Earth. The hyperbolic PDE of interest relates the seismograms, the time-resolved records of ground

motion sampled across the measurement surface $\delta\Omega_{\text{top}}$, to the inferred sources $f(\mathbf{r}, t)$ producing the wave motion and the internal wave-speed distribution, $c(\mathbf{r})$ that we seek to recover. Limiting ourselves to the acoustic case (Tarantola 1984a), the primary observable is pressure, $u(\mathbf{r}, t)$, which satisfies the suitably initialized acoustic wave equation

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = f \qquad \text{within } \Omega, \tag{9}$$

$$u = u_{\text{obs}} \qquad \text{on } \delta\Omega_{\text{top}}. \tag{10}$$

The unknown wave speed c is in $\text{m}\,\text{s}^{-1}$, the acoustic pressure u is in $\text{kg}\,\text{m}^{-1}\,\text{s}^{-2}$. Obtaining the three-dimensional wave-speed $c(\mathbf{r})$ from boundary observations of the wavefield $u(c)$ through the Navier–Cauchy–Hooke mapping, Eq. (9), exemplifies the mathematical problem of seismological inversion that remains of interest to (geo)mathematicians (Freeden 2015; Stefanov et al. 2019). The complex geometry of the Earth's interior, the rich nature of the seismic wavefield, and the diverse regimes under which it can—or cannot—be observed have led to as many solution strategies (Akçelik et al. 2002; de Hoop et al. 2009; Nolet 2015; Symes 2009; Tromp 2020). Lastly, we note the coupled nature of the seismic and the gravitational inverse problems (Berkel and Michel 2010; Liu and Tromp 2009; Michel 2015). For the purpose of this paper, it suffices to note that the acoustic wave speed c relates to the mass density ρ via $c = \sqrt{\kappa/\rho}$ through κ, the bulk modulus.

As in the gravimetric inverse problem, we restrict the observations to discrete receiver locations on the top surface $\delta\Omega_{\text{top}}$. We consider the source term f known (in practice, we use a second-derivative-of-the-Gaussian, or "Ricker" source wavelet, $f(t) = 0.2e^{-1000(t-0.07)^2}$), and we neglect to write the absorbing boundary conditions to simplify the exposition.

We minimize the quadratic misfit function, subscripted H for "Hooke",

$$F_H(u) = \frac{1}{2}\int_0^T \int_{\Omega_{\text{obs}}} \left[u_{\text{obs}} - u(c)\right]^2 dx\, dt, \tag{11}$$

over the spatial boundary dimension and integrated over the time window of interest, $0 \leq t \leq T$, i.e. for a finite, bandlimited "arrival" pulse. When the time-window is long and the bandwidth is large, misfit functions of the type in Eq. (11) are known as "full-waveform". In order to use the wave equation in Eq. (9) as a constraint in the minimization, we write the full Lagrangian of the system,

$$\begin{aligned} \mathscr{L}_H = &\frac{1}{2}\int_0^T \int_{\Omega_{\text{obs}}} \left[u_{\text{obs}} - u\right]^2 dx\, dt \\ &- \int_0^T \int_\Omega \left[\frac{1}{c^2}\frac{\partial u}{\partial t}\frac{\partial v}{\partial t} - \nabla u \cdot \nabla v\right] dx\, dt - \int_0^T \int_\Omega f\, dx\, dt. \end{aligned} \tag{12}$$

The first variation of Eq. (12) with respect to the appropriate test function v and its gradient ∇v results in the weak form of Eq. (9).

The second variation, with respect to the solution u, returns the adjoint equation:

$$\frac{\partial \mathscr{L}_H}{\partial u}(\hat{u}) = \int_0^T \int_{\Omega_{\text{obs}}} \left[u_{\text{obs}} - u(c) \right] \hat{u} \, dx \, dt$$
$$- \int_0^T \int_\Omega \left[\frac{1}{c^2} \frac{\partial v}{\partial t} \frac{\partial \hat{u}}{\partial t} - \nabla v \cdot \nabla \hat{u} \right] dx \, dt - \int_\Omega \frac{\partial v}{\partial t}(T) \frac{\partial \hat{u}}{\partial t}(T) \, dx = 0, \tag{13}$$

to be solved for v. Note that the right-hand-side of Eq. (13) is only known at $t = T$, in contrast to the forward problem where the source evolves from $t = 0$ to $t = T$. This final-value problem thus has to be solved backwards in time from $t = T$ to $t = 0$. Again, the interpretation of the adjoint equation is that it solves the original wave equation, backwards in time and with updated boundary conditions, but now sourced by the waveform differences between the observations and the forward fields.

The third variation, with respect to the design variable c, results in the weak form of the gradient of the misfit function, namely

$$\frac{\partial \mathscr{L}_H}{\partial c}(\hat{c}) = - \int_\Omega \frac{2}{c^3} \left[\int_0^T \frac{\partial u}{\partial t} \frac{\partial v}{\partial t} \, dt \right] \hat{c} \, dx = \frac{d F_H}{dc}. \tag{14}$$

Here, v is the solution of the adjoint equation Eq. (13), propagating in time reversal, and u is the solution of the forward problem, propagated forward in time. Thus, Eq. (14) is the convolution of the forward and the adjoint solutions, a weighted interaction integrated against the partial derivative of Eq. (9) with respect to c.

The above derivations are merely rudimentary sketches of material extensively treated by Tarantola (1984a) and Tromp et al. (2005) using Green function approaches under the Born approximation, by Liu and Tromp (2006) and Liu and Tromp (2009) via the method of Lagrange multipliers, and by Fichtner et al. (2006b), Fichtner et al. (2006a), Plessix (2006), and Virieux and Operto (2009) in an abstract operator formalism.

It has been clearly shown over the last few decades (e.g. Tromp 2020; Virieux and Operto 2009) that the PDE-constrained minimization approach known as full-waveform inversion (FWI) is capable of rather precisely imaging Earth's interior and subsurface structures. While the potential for geologic interpretation increases when gravitational and seismic observables are being treated in tandem, the joint resolution of density and wave speed anomalies is generally difficult with the types of measurements typically at hand. Fortunately, we have recourse to a last type of observables and a final, nonlinear PDE with which to constrain data misfit functions, as we shall see in the next section.

2.3 Geodynamics

The Earth's silicate mantle behaves as a highly viscous fluid over geologic timescales (Ranalli 1995; Schubert et al. 2001), and the description of its structure as an incompressible "Stokes" flow at steady state (Kennett and Bunge 2008; Malvern 1969) is a

useful abstraction. The velocity $\mathbf{u}(\mathbf{r})$ and stress $\boldsymbol{\sigma}(\mathbf{r})$ fields are governed by mass and momentum conservation equations with free-slip boundary conditions on both sides and the bottom, and a zero-traction boundary condition at the free-surface top:

$$\nabla \cdot \mathbf{u} = 0 \quad \text{within } \Omega, \qquad \mathbf{u} \cdot \hat{\boldsymbol{\gamma}} = 0 \quad \text{on } \delta\Omega \backslash \delta\Omega_{\text{top}}, \tag{15}$$

$$\nabla \cdot \boldsymbol{\sigma} = -\rho\mathbf{g} \quad \text{within } \Omega, \qquad \boldsymbol{\sigma} \cdot \hat{\boldsymbol{\gamma}} = \mathbf{0} \quad \text{on } \delta\Omega_{\text{top}}. \tag{16}$$

Here, ρ is the mass density, $\hat{\boldsymbol{\gamma}}$ is the outward normal vector, and the constant gravitational acceleration $\mathbf{g}$. Stokes' constitutive law relates the total stress $\boldsymbol{\sigma}$ to the flow-induced strain rate $\dot{\boldsymbol{\varepsilon}}$ as follows:

$$\boldsymbol{\sigma} = -p\mathbf{I} + 2\eta\dot{\boldsymbol{\varepsilon}}, \qquad \dot{\boldsymbol{\varepsilon}} = \frac{1}{2}\left(\nabla\mathbf{u} + [\nabla\mathbf{u}]^{\text{T}}\right), \tag{17}$$

where $p(\mathbf{r})$ is the static equilibrium pressure of the fluid at rest, $\mathbf{I}$ the identity tensor, and η the (effective) dynamic viscosity, constant for Newtonian flows, or, allowing for the explanation of a fuller range of observable Earth phenomena both in the lab and at the plate-tectonic scale (Malevsky and Yuen 1992; Parmentier et al. 1976), for the case of power-law creep (Schubert et al. 2001), the strain-rate dependent

$$\eta(\dot{\boldsymbol{\varepsilon}}) = \eta_0 \left(\frac{\dot{\varepsilon}_{II}}{\dot{\varepsilon}_0}\right)^{\frac{1}{n}-1}, \tag{18}$$

where η_0 and $\dot{\varepsilon}_0$ are reference values, $\dot{\varepsilon}_{II} = (1/2\dot{\boldsymbol{\varepsilon}} : \dot{\boldsymbol{\varepsilon}})^{1/2}$ is the second invariant of the strain rate, and $1 \lesssim n \lesssim 5$ the applicable power-law coefficient (Petra et al. 2012; van den Berg et al. 1993).

The velocity $\mathbf{u}(\mathbf{r})$, in m s^{-1}, and the pressure $p(\mathbf{r})$, in Pa, are the solutions to the PDE system (15)–(17), or indeed, the full equation of motion

$$-\nabla p + \rho\mathbf{g} + \nabla \cdot \left[2\eta(\dot{\boldsymbol{\varepsilon}})\dot{\boldsymbol{\varepsilon}}\right] = \mathbf{0}. \tag{19}$$

The unknown parameters of interest are ρ, the mass density in kg m^{-3}, the reference viscosity η_0, in Pa s, and n, the dimensionless stress exponent.

In order to define the Lagrangian for the misfit function, subscripted S for "Stokes",

$$F_S(\mathbf{u}) = \frac{1}{2} \int_{\Omega_{\text{obs}}} [\mathbf{u}_{\text{obs}} - \mathbf{u}(\rho, \eta_0, n)]^2 \, dx, \tag{20}$$

which we seek to minimize with respect to the PDE constraint Eq. (19), we combine the solutions of Eq. (15) into a single solution $(\mathbf{u}, p)$. The full Lagrangian reads:

$$\begin{aligned}
\mathcal{L}_S = \frac{1}{2} \int_{\Omega_{\text{obs}}} [\mathbf{u}_{\text{obs}} - \mathbf{u}]^2 dx + \int_{\Omega} 2\eta(\dot{\boldsymbol{\varepsilon}})\dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, dx \\
- \int_{\Omega} p \, \nabla \cdot \mathbf{v} \, dx - \int_{\Omega} (\nabla \cdot \mathbf{u}) \, q \, dx + \int_{\Omega} \rho\mathbf{g} \cdot \mathbf{v} \, dx.
\end{aligned} \tag{21}$$

The first variation of the Lagrangian with respect to an appropriate set of test functions $(\mathbf{v}, q)$ results in the weak form of Eq. (19).

The second variation, with respect to the solutions $(\mathbf{u}, p)$ in some new direction $(\hat{\mathbf{u}}, \hat{p})$ results in the adjoint equation:

$$\frac{\partial \mathscr{L}_S}{\partial (\mathbf{u}, p)}(\hat{\mathbf{u}}, \hat{p}) = \int_{\Omega_{\mathrm{obs}}} \left[\mathbf{u}_{\mathrm{obs}} - \mathbf{u}\right] \hat{\mathbf{u}} \, dx + \int_\Omega \mathbb{D}(\mathbf{u})\dot{\boldsymbol{\varepsilon}}(\mathbf{v}) : \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) \, dx$$
$$- \int_\Omega (\boldsymbol{\nabla} \cdot \mathbf{v})\hat{p} \, dx - \int_\Omega q \, \boldsymbol{\nabla} \cdot \hat{\mathbf{u}} \, dx = 0. \tag{22}$$

In this equation $(\mathbf{u}, p)$ are the forward velocity and pressure, respectively. The equation is to be solved for the adjoint solution $(\mathbf{v}, q)$ using appropriate test functions $(\hat{\mathbf{u}}, \hat{p})$. $\mathbb{D}(\mathbf{u})$ is the derivative of the nonlinear viscosity with respect to the forward velocity (e.g Reuber et al. 2020):

$$\mathbb{D}(\mathbf{u}) = 2\eta(\dot{\boldsymbol{\varepsilon}}) \left(\mathbb{I} + \left(\frac{1}{n} - 1\right) \frac{\dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \otimes \dot{\boldsymbol{\varepsilon}}(\mathbf{u})}{2\dot{\varepsilon}_{II}^2} \right), \tag{23}$$

with $\mathbb{I}$ being the fourth-order identity tensor. Note that compared to the PDEs in the previous sections, Eq. (22) is not self-adjoint due to the nonlinear viscosity (18).

The third variation of the Lagrangian with respect to the design parameter set $\mathbf{m} = \begin{pmatrix} \rho & \eta_0 & n \end{pmatrix}^{\mathrm{T}}$ in the new direction $\hat{\mathbf{m}} = \begin{pmatrix} \hat{\rho} & \hat{\eta}_0 & \hat{n} \end{pmatrix}^{\mathrm{T}}$ reads:

$$\frac{\partial \mathscr{L}_S}{\partial \mathbf{m}}(\hat{\mathbf{m}}) = \begin{pmatrix} \int_\Omega \mathbf{g} \cdot \mathbf{v} \, \hat{\rho} \, dx \\ \int_\Omega 2 \, (\dot{\varepsilon}_{II}/\dot{\varepsilon}_0)^{\frac{1}{n} - 1} \, \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, \hat{\eta}_0 \, dx \\ - \int_\Omega \frac{\eta(\dot{\boldsymbol{\varepsilon}}) \log(\dot{\varepsilon}_{II}/\dot{\varepsilon}_0)}{n^2} \, \hat{n} \, dx \end{pmatrix} = \frac{dF_S}{d\mathbf{m}}, \tag{24}$$

where $\mathbf{u}$ and $\mathbf{v}$ are the forward and adjoint velocity fields, respectively, and p and q are the forward and adjoint pressure terms.

The adjoint Stokes method has been firmly established for geodynamic inverse problems (Bunge et al. 2003; Ghelichkhan and Bunge 2016; Horbach et al. 2014), revealing, e.g., mantle structure from plate-motion histories (Bunge et al. 2002), plate-boundary coupling (Ratnaswamy et al. 2015), the thermal conditions of the mantle (Colli et al. 2018), and rheological parameters in steady-state models (Reuber et al. 2018).

2.4 Combining cost functions

In this work we aim to minimize the combined cost function

$$F = F_N(u) + F_H(u) + F_S(\mathbf{u}). \tag{25}$$

We note that this function and its gradients are additive, leading to an optimization procedure that is uncoupled in theory, and the different parameter updates and the forward problem solutions do not mutually influence each other. Recent studies (Crestel

et al. 2018) have suggested introducing an explicit coupling via, e.g. cross-gradient or density-norm regularization. Here we focus solely on the algorithmic structure of the uncoupled inverse problem. However, cross-product regularization can be approximated by updating the parameters only at the points for which all gradients agree in the update direction, as we shall show.

3 Geological settings

As we wrote in Sect. 1.3 we consider three geological settings in this work: highly idealized end-members that nevertheless comprise recognizable scenarios with considerable complexity. Relevant material parameters will be summarized in Table 1.

3.1 Falling block (FB)

The first scenario is of a "falling" block (FB) of a certain mass and viscosity, sinking in a less dense and less viscous matrix. Such a setting may occur on many different scales in nature, for instance when crystals settle in a magma chamber (Martin and Nokes 1988, 1989), or when a portion of Earth's lithosphere "delaminates" (Elkins-Tanton 2005, 2007). The geometry is shown in Fig. 1. In our scenario the block has a higher seismic wave speed. The synthetic data are produced with the viscosity, wave speed and density distributions shown.

3.2 Subduction zone (SZ)

The second setting is that of a subduction zone (SZ), a bent descending "slab" (Melosh and Raefsky 1980; Sleep 1975): colder, denser, and more viscous than the ambient mantle, and with a higher seismic wave speed. The geometry is shown in Fig. 2, as are the data synthetics for the gravity anomaly, the surface velocity, and the seismograms generated by sources at depth, recorded by receivers at the surface.

Table 1 Material parameters used in the different geological scenarios

Scenario	ρ (AP/BG)	c (AP/BG)	η_0 (AP/BG)	n (AP/BG)
Falling block (FB)	1.5/1.0	1.4/1	1000/1	1/1
Subduction zone (SZ)	1.5/1.2	1.4/1	1000/1	1/1
Mantle plume (MP)	1.0/1.2	1/0.8	10/10	1/1

Here, ρ is mass density (as in Sects. 2.1–2.3), c is acoustic wave speed from Sect. 2.2), and η_0 is the reference viscosity and n the power-law coefficient from Sect. 2.3

FB falling block, *SZ* subduction zone, *MP* mantle plume, *BG* background material parameter, *AP* anomalous material parameter
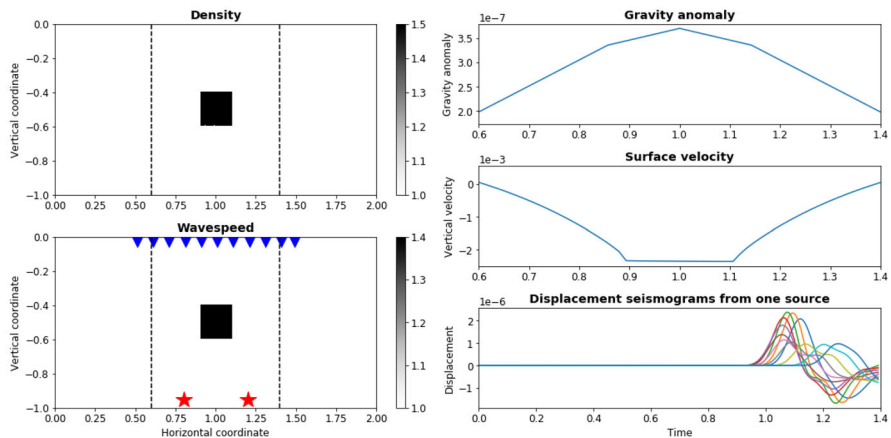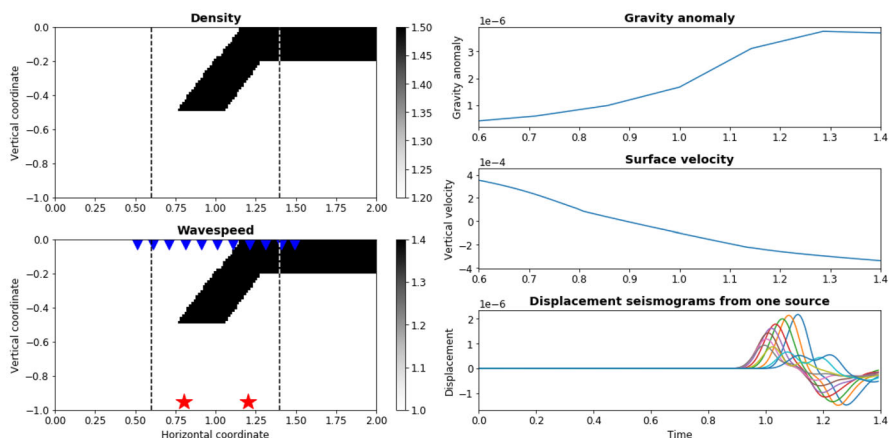
Fig. 1 Geometry of the falling block (FB) scenario (left column) and computed geophysical observables (right column). Upper-left panel shows the density distribution. Lower-left panel shows the seismic wave speed distribution, and the locations of sources at depth (red stars) and surface receivers (blue triangles). Upper-right panel shows the observable gravity anomaly, middle right the vertical surface velocity, and lower right the displacement seismograms from one of the sources to each of the receivers (color figure online)

Fig. 2 Geometry of the subduction zone (SZ) scenario (left column) and computed geophysical observables (right column). Upper-left panel shows the density, lower-left panel the wave speed, and the source-station geometry. Right columns show the geophysical observables, in the same layout as Fig. 1

3.3 Mantle plume (MP)

The third setting is intended to capture the essentials of a mantle plume: a rising upwelling, hotter than, and as such with lower density, wave speed and viscosity than, the ambient mantle (Morgan 1971; Sleep 1990; Wilson 1973). As presented in Fig. 3, we abstract this setting by adding a Gaussian shaped perturbation at the bottom of the modelling domain. A rigid crust is located at the top.
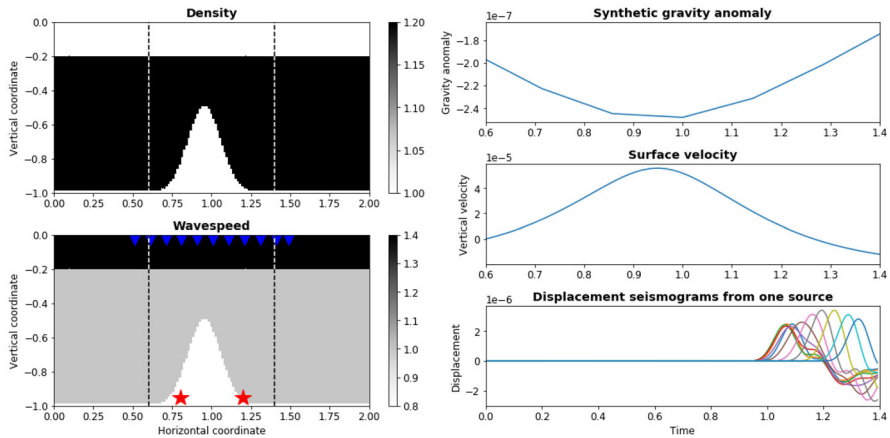
Fig. 3 Geometry of the mantle plume (MP) scenario (left column) and synthetic geophysical observables (right column). As in Figs. 1 and 2 the upper left shows the density distribution, lower left panel the wave speed distribution. Panels in the right columns show, from top to bottom, the gravity anomaly, vertical surface velocity, and displacement seismograms at different receivers

4 Numerical experiments

In this section we present the collection of numerical results. We first formulate a motivation based on grid-search sampling, highlighting that a combined cost function indeed contains more information than any of them separately. This is followed by a visualization of the "kernels" of the three sets of governing PDEs. These kernels are the derivatives of the forward solution with respect to the material parameters. We present examples for continuous-field recovery using pointwise adjoint gradients subject to the three types of observable constraints, both sequentially and iteratively. Finally, we propose a geometric update algorithm based on the pointwise gradients.

4.1 Shape of the cost function

At first, one can show that the coupled cost function of Eq. (25) is indeed "better" than the respective single cost functions. With "better" we aim for (i) a more uniquely defined, and (ii) a steeper cost function, which may increase the convergence rate of the optimization procedure.

A straightforward approach involves performing a grid search. A grid search algorithm only requires discretization of the forward model over the parameter space and an evaluation of the forward model at each discrete point. We choose the viscosity, the density, the wave speed, and the depth to the tip or center of the anomaly as parameters. Wave speed and density are coupled such that an increase in density implies an increase in wave speed. We discretize the parameter space as a $10 \times 10 \times 3$ viscosity-density-depth matrix. The geometry of the structure (apart from the tip) is considered known. The synthetic data are always computed with the parameter combination that is exactly in the center of the discretized parameter space.

In order to compare the cost functions we normalize them as

$$F_\beta^* = \frac{F_\beta - \bar{F}_\beta}{\text{std}(F_\beta)} - \min\left[\frac{F_\beta - \bar{F}_\beta}{\text{std}(F_\beta)}\right], \tag{26}$$

where F_β is an individual cost function, i.e. F_N for the gravitational (Newton) problem (Eq. 3), F_H for the seismoacoustic (Hooke) problem (Eq. 11) and F_S for the geodynamics (Stokes) problem (Eq. 20), $\bar{F}_\beta$ denotes the mean of all such cost functions calculated during the grid search and std denotes the standard deviation. We recall that F denotes the combined cost function of Eq. (25).

All three geological settings are investigated, starting with the falling block (FB) example scenario introduced in Fig. 1. In Fig. 4 the size and color of the circles indicate how good (small, light gray) or bad (large, dark gray) the fit is at the considered point. The top two panels show the influence of the viscosity of the falling block. In this particular setup it appears to be almost irrelevant. The three separate cost functions shown in the second row all show a "valley" of good solutions, which illustrates the difficulty for simple gradient-descent algorithms to converge.

While curvature information from the Hessian matrix could improve convergence of such problems, the different shapes of the valleys suggest that a combination of cost functions might be better behaved. The final combined cost function F is shown in the bottom panel: it indeed satisfies our notion of being 'better', i.e., steeper and more unique than each cost function taken separately.

The second test case is the subduction zone (SZ) example of Fig. 2. The top two panels in Fig. 5 suggest that the viscosity of the subducting part does have an impact on the cost function shape, which is, however, much smaller than the equivalent one for the density. Again, the combined cost function, shown in the bottom panel, is essentially unique, and much steeper than any separate ones. One important note is that a change in the subduction depth also changes the overall volume of material. This relation has implications for the gravitational anomaly which is mainly driven by the volume and amplitude of the anomalous material present at depth.

The third test case is the mantle plume (MP) scenario of Fig. 3, with the results presented in Fig. 6. As in the case before, a change in the tip of the Gaussian anomaly also results in a change of the overall volume of material. The impact of viscosity is limited and only pronounced in the limit that the plume is significantly more viscous than the surrounding mantle, which is geodynamically implausible. In this scenario even the combined cost function shows an undesirable valley of acceptable solutions, but it is steeper than the separate cost functions.

4.2 Continuous field recovery by discretized inversion

Section 4.1 showed how using a combined cost function helps in defining a better-posed minimization problem than using cost functions separately. The question remains how a non-grid-search deterministic inversion approach might make use of this beneficial cost surface shape. For the purpose of continuous-field recovery or kernel calculation, the parameter space spans the discretized range of the forward problems, substantially
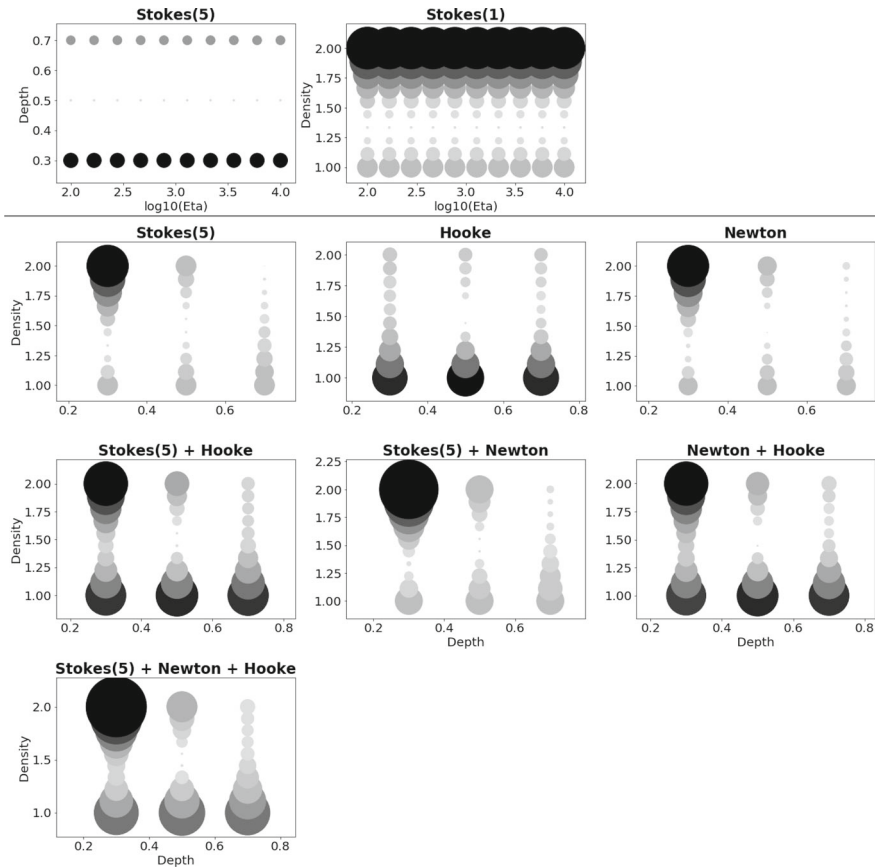
Fig. 4 Sampled cost functions for the falling block (FB) scenario. The darker and the larger the circles the higher the value of the cost function at the sample point. Upper row: cost function for the surface velocity, for depth versus viscosity, and density versus viscosity, respectively. Here, and in the following, Stokes(i) refers to the value of the respective index i of the third and fixed variable for the Stokes equation, i.e. Stokes(5) refers to a viscosity value of 1000 or a density value of 1.5 while Stokes(1) refers to a depth value of 0.5. As supported by the Stokes sinker theory, the viscosity of the falling block exerts no influence on the surface velocity. All axes below the line are in terms of the depth of the anomaly against the primary design variables density or wave speed. We label wave speed, ranging from 0.9 to 1.9, after scaling to density. The second row shows the cost function for the surface velocity (Stokes), the seismic wave displacement (Hooke) and the gravitational potential (Newton). The third row shows possible combinations of the pure cost functions. The bottom panel shows the combined cost function for all observables. The bottom-row panel shows that the combined cost function is indeed substantially steeper than each separate cost function and will lead to more unique solutions than is afforded by, for instance, minimization for the gravitational cost-function by itself

increasing the dimensionality. We seek the pointwise parameter field, based on the pointwise adjoint gradients, after assuming a homogeneous initial guess, and ignoring any prior information on space, time, or the parameter of interest.
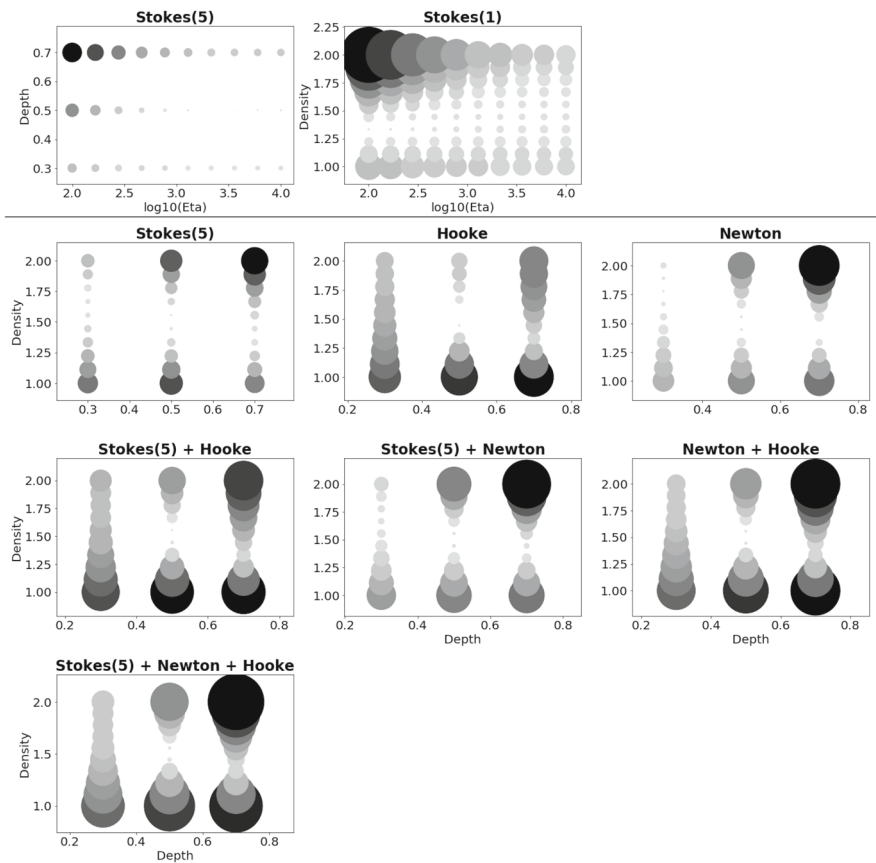
Fig. 5 Sampled cost functions for the subduction zone (SZ) scenario. Axes, labels and plot arrangement are as in Fig. 4. Density-equivalent seismic wave speeds again range from 0.9 to 1.9. Upper row panels: cost function for the surface velocity, for depth versus viscosity, and density versus viscosity, respectively. Second row: cost function for the surface velocity (Stokes), the seismic displacement field (Hooke) and the gravitational potential (Newton). Third row shows combinations of the pure cost functions. The bottom-row panel shows the cost function combined for all observables, highlighting its benefits

4.2.1 Sensitivity kernels

To start out, it is useful to inspect the sensitivity of the forward solutions to the pointwise material parameters that control them. This sensitivity kernel for material parameters will have a value proportional to their influence on the forward solution, subject to the current initial configuration. As such, it affords insight into the information content of the purely forward solutions in terms of being able to solve for the material parameters as part of the inverse problem.

The sensitivity kernel of each of the PDEs contained in Eqs. (1)–(2), (9)–(10), and (15)–(17), can be derived by assuming a cost function equal to the forward solution, i.e. by changing the expressions F_β in Eqs. (3), (11) and (20) to
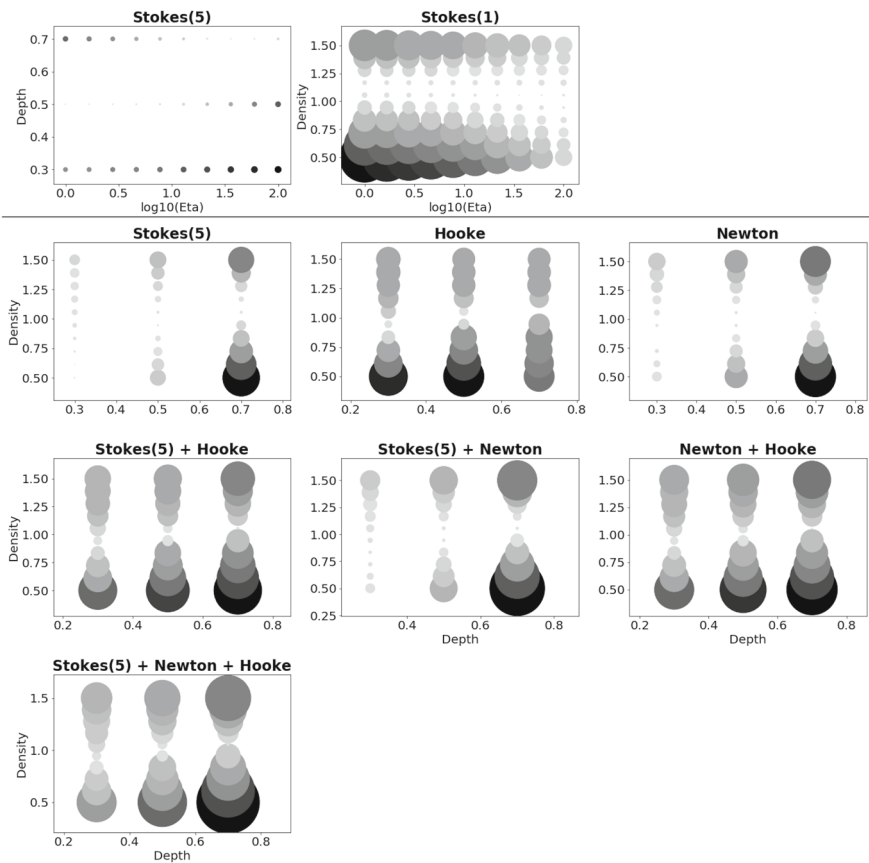
Fig. 6 Sampled cost functions for the mantle plume (MP) scenario. Axes, labels and plot arrangement are again as in Figs. 4 and 5, but density-corresponding values of the seismic wave speed range from 0.3 to 1.3. Upper row: cost functions for the surface velocity for depth versus viscosity, and density versus viscosity. Second row: the cost function for the surface velocity (Stokes), the seismic displacement (Hooke) and the gravitational potential (Newton). The third row shows combinations of the pure cost functions. The bottom-row plot shows the combined cost function for all observables. The combined cost function is substantially steeper than each of the separate cost functions, and more unique than, for instance, the gravitational-potential cost function by itself. However, even in this case the combined cost function remains non-unique

$$\hat{F}_\beta = \int_{\Omega_{\text{obs}}} u(\mathbf{m}) \, dx, \qquad (27)$$

whereby $\mathbf{m}$ is the parameter (set) of interest, i.e. mass density ρ for the Newton problem, acoustic wave speed c for the Hooke problem, and (ρ, η_0, n) for the Stokes problem, where η_0 and n are the reference viscosity and the power-law exponent, respectively. Following the same framework as described in Sect. 1 we obtain, instead of the gradient of the cost function, the gradient of the forward operator on the boundary Ω_{obs}, with respect to the parameters $\mathbf{m}$. Only the adjoint source, that is, the right-hand-side of the adjoint equations Eqs. (5), (13), and (22), changes.

For the example of the gravitational potential the adjoint equation now reads:

$$\frac{\partial \mathcal{L}_N}{\partial u}(\hat{u}) = \int_{\Omega_{\text{obs}}} \frac{\partial \hat{F}_N}{\partial u} \hat{u} \, dx - \int_{\Omega} \nabla v \cdot \nabla \hat{u} \, dx = 0, \tag{28}$$

noting that $\partial \hat{F}_N / \partial u = 1$, and *mutatis mutandis* for the other PDEs.

The gradient equations remain unchanged—a change in the cost function, without regularization, only affects the adjoint source—and one retrieves (i) $d\hat{F}_N/d\rho$ the sensitivity of the gravitational potential to the density, (ii) $d\hat{F}_H/dc$, the sensitivity of the displacement to the wave speed, and (iii) $d\hat{F}_S/d\mathbf{m}$, the sensitivity of the vertical surface velocity to the density, viscosity or power-law exponent, respectively.

Since the kernels image the sensitivity of the gravitational potential, the displacement due to an acoustic wave, and the surface velocity, with respect to density, wave speed, and viscosity, respectively, they are to be interpreted as resolution indicators. Where the kernels vanish the material parameter has no discernible effect on the observations, and any updates of or interpretations based on recovered anomalies will be spurious. It should nevertheless be noted that all kernels change shape in function of their initial conditions, and thus are only valid for the current state. The kernels for the falling block scenario are summarized in Fig. 7, the ones for the subduction zone setting in Fig. 8 and those for the mantle plume in Fig. 9.

The kernel of the gravitational potential with respect to the density subject to the Poisson equation is very local to the measurement points followed by a diffuse profile towards the bottom of the domain. As such this measurable is also not very useful for imaging a subsurface structure, as is well known in the literature.

The wave speed kernel shows a complicated pattern due to reflections at the interfaces and its time history. Systematic study of the kernels for wave speed anomalies is the subject of extensive literature in the seismological community (Tromp et al. 2005).

The kernel for the surface velocity with respect to the density subject to the Stokes equations has a very diffuse pattern. Almost all density nodes have an effect on the surface velocity. This quantity may thus not be very useful to reconstruct the geometry of an anomaly. In comparison, the viscosity kernel is very localized around the anomaly, suggesting that even a small change in the viscosity around the anomaly will influence the surface velocity.

4.2.2 Collinear gradient updates

As we have seen, different governing equations reveal different features. The uppermost row in Fig. 10 shows the pointwise gradient for the falling block scenario, for the density subject to the Stokes equations ($dF_S/d\rho$), the viscosity in the Stokes equations ($dF_S/d\eta$), the wave speed from the wave equation (dF_H/dc), and the density from the gravitational potential ($dF_N/d\rho$). Only the wave-speed gradient is localized enough to reveal the geometry of the anomaly. The two gradients from the Stokes equation are rather diffuse, as is that of the Poisson equation. The same behavior can be seen for the two other cases—the subduction zone scenario shown in the top rows of Fig. 11, and for the mantle plume scenario shown in the top row of Fig. 12.
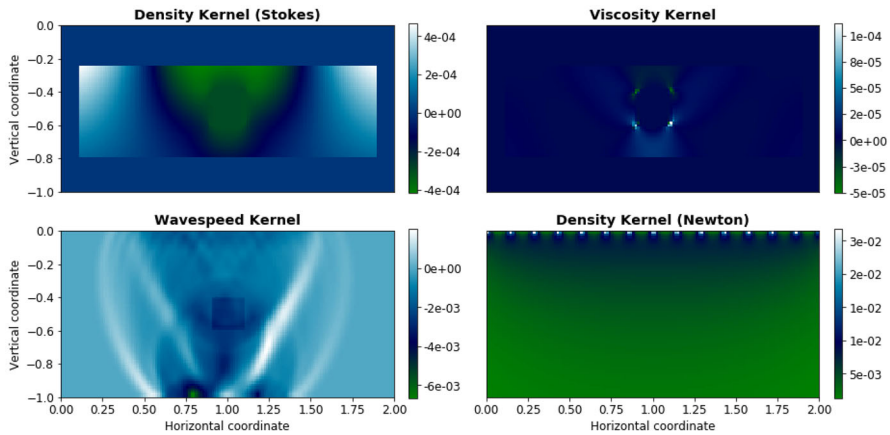
Fig. 7 Sensitivity kernels for the Stokes flow, Hookean wavefield, and Poisson equations in the falling block scenario. The upper left panel shows the kernel for the surface velocity with respect to the density. The kernel is very diffuse and suggests that the block has the highest impact on the surface velocities. The upper right panel shows the kernel of the surface velocity with respect to the viscosity. This kernel is very localized to the edges of the block, suggesting that changing the viscosity at these points will most affect the surface velocity. The lower left panel shows the kernel of the wave speed with respect to the seismic deformation. This shows a predictably high influence along the ray paths. The lower right plot shows the kernel of the density with respect to the gravitational potential. Due to the nature of the Poisson equation this kernel simply shows a diffuse pattern decreasing with depth
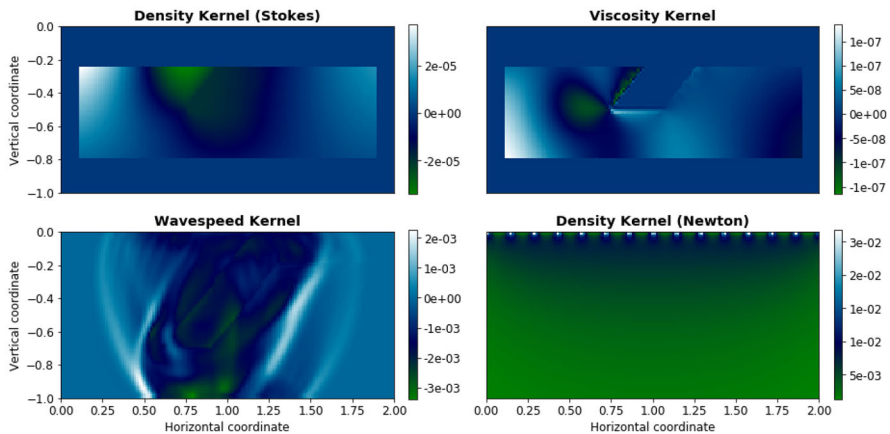
Fig. 8 Sensitivity kernels for the wavefield, flow and gravity equations for the subduction zone scenario, laid out identically to Fig. 7. The kernel for the surface velocity with respect to the density (upper left) is again very diffuse and suggests that the block has the highest impact on the surface velocities. The kernel of the surface velocity with respect to the viscosity (upper right) is very localized to the edges of the block, suggesting that changing the viscosity at these points will affect the surface velocity the most. The wave speed kernel with respect to the seismic deformation (lower left) shows a high influence along the ray paths. The lower right plot shows the kernel of the density with respect to the gravitational potential. Due to the nature of the Poisson equation this kernel again simply diffuses away with increasing depth
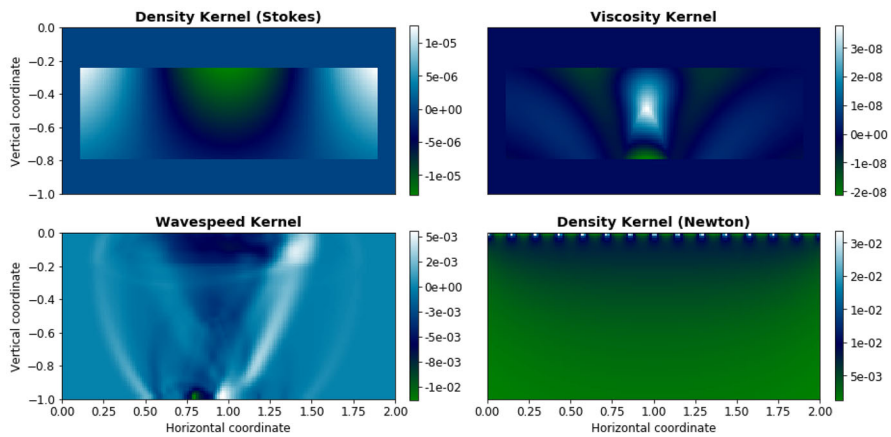
Fig. 9 Sensitivity kernels for the wavefield, Stokes and Poisson equations for the mantle plume scenario, laid out as in Figs. 7 and 8. The surface velocity with respect to density kernel (upper left) is diffuse and indicates that the block has the highest impact on the surface velocities. The kernel of the surface velocity with respect to the viscosity (upper right) does not show any localized structure: there is no difference in viscosity for this scenario. The lower left panel with the kernel of the wave speed with respect to the seismic deformation shows a strong influence concentrated along the seismic ray paths. The lower right plot shows the kernel of the density with respect to the gravitational potential. Due to the nature of the Poisson equation this kernel mainly shows a diffuse pattern decreasing with depth

The second, third and fourth rows in Figs. 10, 11 and 12 show the areas where the respective sum of gradients (normalized by their standard deviation, as in Eq. 26) point in the same direction. Nodes in which the gradients would steer the solution toward different directions should probably not be considered in an inversion framework: updates there will necessarily result in a worse update of the cost function for any one of the observables. By design we consider that an increase in the viscosity has to mimic an increase in the density and the wave speed, and vice versa. The same behavior can also be achieved by adding a regularization that opts to optimize the cross product of the gradients, by imposing a penalty on points where the gradients point in different directions—an approach that is described in more detail elsewhere (e.g., Crestel et al. 2018).

As expected, only considering Stokes flow and the gravitational potential will not improve the solutions much. However, the combination of all available gradients is substantially closer to the synthetic shape than the one derived from the gradients of the Stokes and gravitational potential, respectively, and even better than the wave speed gradient by itself.

4.2.3 Sequential gradient updates

After investigating the sensitivity kernels and initial cost-function gradients, the latter can be used for gradient-descent inversions. The intuitive approach is to update the parameters sequentially. This approach is in common use, since community codes usually only optimize one of the cost functions for either wave speed, density based on gravitational potential or viscosity and density based on flow. This workflow is
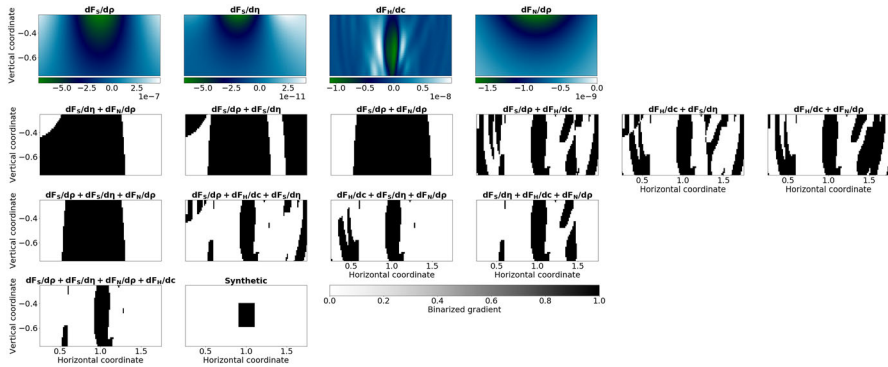
Fig. 10 Cost-function gradients for the falling block scenario. The upper row shows the raw gradients for the surface velocity with respect to the density ($dF_S/d\rho$) and the viscosity ($dF_S/d\eta$), the deformation caused by the seismic wave with respect to the wave speed (dF_H/dc) and the gravitational potential with respect to the density ($dF_N/d\rho$). In the panels below the gradients are quantized, such that only nonzero values and collinear gradients are emphasized. The second row shows first-order combinations of the quantized raw gradients, as indicated. The third row shows second-order combinations of those same gradients. The bottom row shows the result of all quantized gradients combined and the target shape of the anomaly. The fully combined gradient—only considering the nodes where all gradients point in the same direction— is substantially closer to the synthetic result than any of the individual raw gradients. The improvement impacts even the wave speed gradient, traditionally a stand-alone tool to invert for anomaly geometry
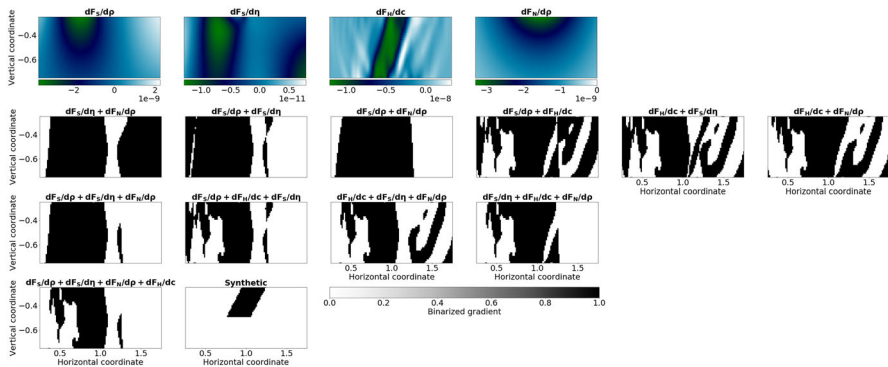
Fig. 11 Cost-function gradients for the subduction zone scenario, in a layout identical to that of Fig. 10 and 11. The top row shows raw gradients. The second row shows first-order combinations of the quantized raw gradients, the third row shows second-order combinations. The last row shows the result of all quantized gradients combined, next to the target shape of the anomaly. The fully combined gradient, when considering the nodes where all gradients point in the same direction, is substantially closer to the synthetic result than any of the raw gradients. The resulting fully combined gradient is broader than the negative part of the raw wave speed gradient. This is due to the fact that the gradient does not discriminate for the sign. The strictly negative part of the gradient is still very close to the synthetic shape, but contains the suggestion for how to change the surrounding background
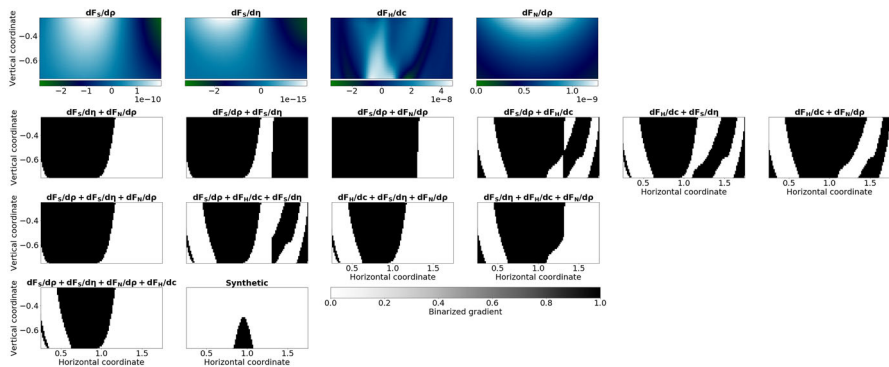
Fig. 12 Cost-function gradients for the mantle plume scenario, laid out as in Figs. 10 and 11. The top row shows the raw gradients, the second row first-order, the third row second-order combinations of the quantized gradients. The last row shows all the quantized gradients combined, and the synthetic shape of the anomaly. The fully combined gradient at the nodes where all gradients point in the same direction is substantially closer to the synthetic result than any of the raw gradients. As in Fig. 11, the fully combined gradient is broader than the negative part of the raw wave speed gradient. This is due to the fact that the gradient does not discriminate for the sign of the anomaly. The purely negative part of the gradient is still very close to the synthetic shape but hints at the need to change the background to reduce the cost function

vulnerable to errors introduced due to post-processing in each separate inversion, or from the unavailability of actual source data. Our work provides the individual pieces to develop one coherent code that inverts for the described observables.

In a sequential framework, one would (i) invert for the pointwise wave speed and thus the geometry of the anomaly, then (ii) choose a contour line in the inverted wave speed, (iii) reduce the number of design parameters by discretizing this contour line and treating other geometries, for instance the crust, as constant phases, (iv) first invert for a density subject to the gravitational potential, which—knowing depth and size of the anomaly—is now unique in a density, and then (v) follow up by inverting for the surface deformation by optimizing the viscosity of the anomaly subject to the Stokes equations. This ultimately leads to a good fit to all observables, and a model that is close to the true solution.

Deriving formal uncertainties on the final solution, which are in part driven by the user's choice of the initial contour line, remains to be explored, e.g. via statistical sampling approaches (Gouveia and Scales 1998) or by investigating the shape of the Hessian (Petra et al. 2014).

Figure 13 presents the sequential workflow for the falling block scenario. From top to bottom, we show the final inversion results for acoustic wave speed (c), mass density (ρ) and viscosity (η) in the panels of the left column, and the behavior of the cost functions for the Hooke (F_H), Newton (F_N) and Stokes (F_S) problems, in descending order, throughout the gradient-descent iterations, normalized by their initial values (F_H^0, F_N^0 and F_S^0). The wave speed optimization, despite being based on only two sources, accurately positions the anomaly in the center of the domain. The inversion for the gravitational potential finds a density contrast of about 1.2, lower than the synthetic value of 1.5, owing to the volume of the recovered anomaly being substantially higher than the target volume. The surface deformation is used to optimize for the viscosity.
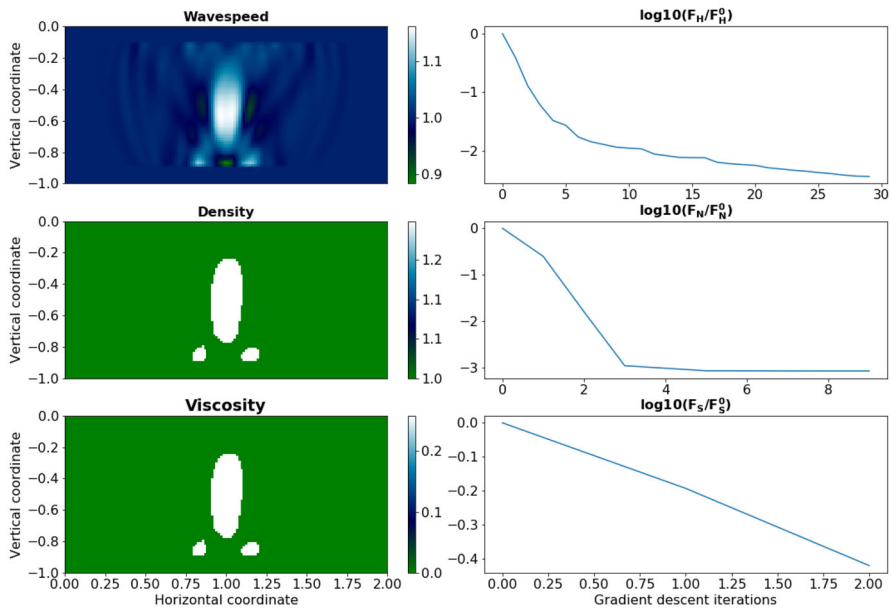
Fig. 13 Summary of the sequential inversion framework for the falling block scenario. The top row shows the inversion of the seismic displacement for the wave speed and the evolution of the cost function with gradient-descent iteration. The middle row shows the converged result and the cost function evolution for the second step of the sequential approach, which inverts the gravitational potential for the density subject to using a geometry for the anomaly chosen on the basis of a contour line of the wave speed inversion result preceding this step. The bottom row shows the final result of the last step of the sequential approach, the inversion for viscosity from the surface deformation. The known true values for neither density nor viscosity were reached. The wave speed contour chosen or the wave speed inversion result itself may not have been accurate enough for a fully successful recovery

Interestingly, this inversion converges on a negative viscosity anomaly, despite the target, shown in Fig. 1, being positive (see Table 1). Figure 14 shows the workflow for the subduction zone scenario. In this case, the seismic inversion almost perfectly captures the shape and size of the wave speed anomaly, which in turn results in a good reproduction of the density and viscosity, as compared with Fig. 2. Figure 15 shows the sequential workflow for the mantle plume scenario. Here, as in the falling block scenario, the outline of the input anomaly could only approximately be imaged from the inversion of the seismic records. As a result, the recovered density is also further away from the synthetic input, see Fig. 3. The flow inversion is omitted here, since the viscosity of the anomaly was unchanged from the background.

4.2.4 Semi-iterative procedure

In contrast to the sequential approach, all parameter fields may receive gradient updates at the same time, either iteratively or by combining gradients of the same parameter. Such an approach may constitute an improvement over the sequential approach from
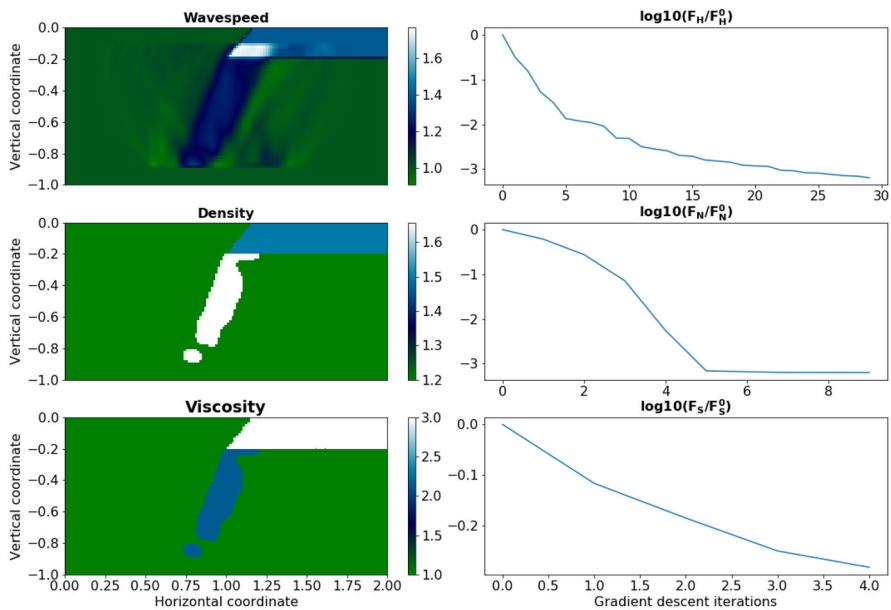
Fig. 14 Sequential inversion framework for the subduction zone scenario, with layout as for Fig. 13. Top: gradient-descent wave speed inversion result, and cost function reduction. Middle: converged result for the second step: density from gravity, within a wave speed contour. Bottom: converged result of the final step: viscosity inverted from surface deformation. Neither density nor viscosity completely match the known input, hinting at the sensitivity of the algorithm to the selection of an appropriate wave speed contour
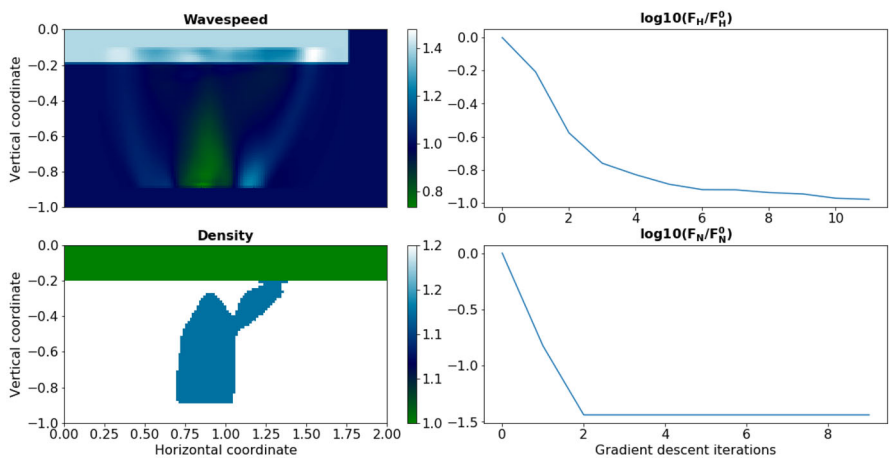
Fig. 15 Summary of the sequential inversion framework for the mantle plume scenario, laid out exactly as Figs. 13 and 14. Top row shows inversion result for the wave speed and the reduction of the cost function. Middle row shows the result of the second step, the converged final result of the inversion for density from the gravitational potential inside of a contour line of the prior wave speed result. Bottom row shows the final result of the third step of the sequential approach, the determination of viscosity from the surface deformation. The density and viscosity are not perfect matches for the known input anomalies
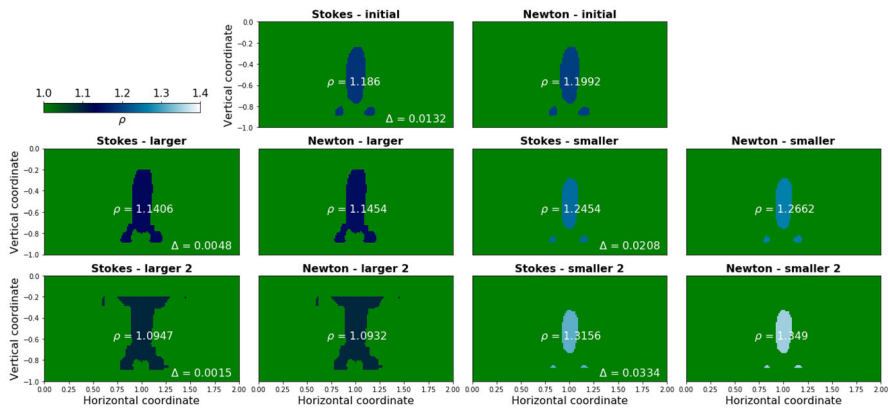
Fig. 16 Summary of the iterative approach for the density inversion under the falling block scenario. Top row shows the results of the inversion for the density with respect to the Stokes field and the Newton potential, respectively, based on an initial binarized subdivision chosen after a contour line of the initial wave speed inversion. The second row shows the pairs of results derived from choosing a slightly larger, or smaller, binarization contour line. The third row shows the results for an even larger and smaller contour line. The inverted density anomalies ρ, constant inside the individual subdomains, are quoted in white. The anomaly difference, between the Stokes and the Newton solutions for each scenario, is labeled Δ

Sect. 4.2.3, as in this approach gradients can immediately react to differences in the flow field caused by parameter updates.

For the mantle plume scenario, adding density gradients of the surface-deformation and the gravitational-potential cost functions, and stopping the algorithm once they diverge, results in a reduction of the combined cost function. However, the main role of such a procedure may well lie in identifying when the initial shape of the target anomaly is wrong. The density inferred from the flow field and from the gravitational potential should agree with each other after the global cost function has been sufficiently reduced to the point where it can be surmised that the anomaly geometry is captured correctly to first order.

One improvement would be to tweak the contour line of the wave speed result that is taken as a basis for continued inversion, depending on the difference between the Stokes and the Newton gradients. The result of such an algorithm is shown in Fig. 16 for the falling block scenario, in Fig. 17 for the subduction zone scenario, and in Fig. 18 for the mantle plume scenario. The density inversion results for the Stokes flow problem and the Newton gravitational potential, respectively, are shown, as is Δ, the difference in density. Once the density difference is minimized, one can finally invert for the corresponding pointwise viscosity subject to the flow solution.

To summarize, this algorithm would start with a pointwise wave speed inversion, followed by a binarization—based on a contour line of the inverted wave speed—of the inverted anomaly in order to avoid non-local effects due to the flow and gravity gradients. Consequently, the algorithm iterates the binarization of the wave speed with a separate update of the density subject to the flow and gravity fields, respectively. Once the difference between the two inverted density fields is minimized, the final viscosity can be inferred from one additional inversion of the flow field.
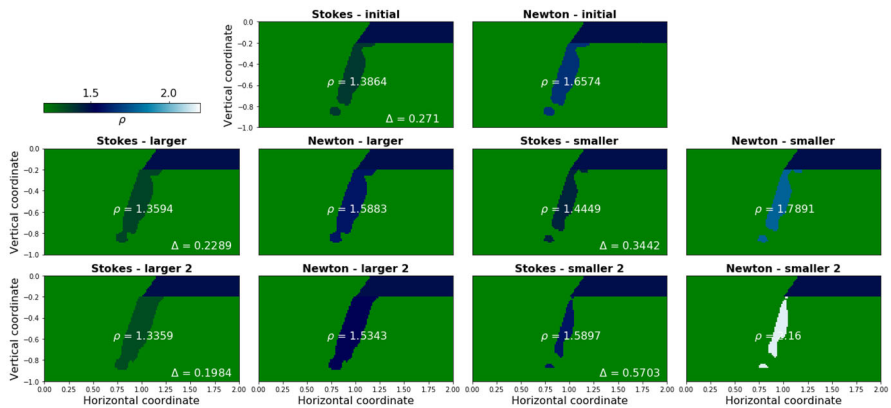
Fig. 17 Iterative approach results for the density in the subduction zone scenario, with a layout and labeling exactly as in Fig. 16. Top row shows density of the Stokes and Newton problems for constant-density subdomains based on a wave speed anomaly contour. The second and third rows show pairs of results based on larger, and smaller, values of the initial wave speed subdivision contour value. Annotations detail recovered density anomalies, ρ, and their differences, Δ, under the different inversion methods
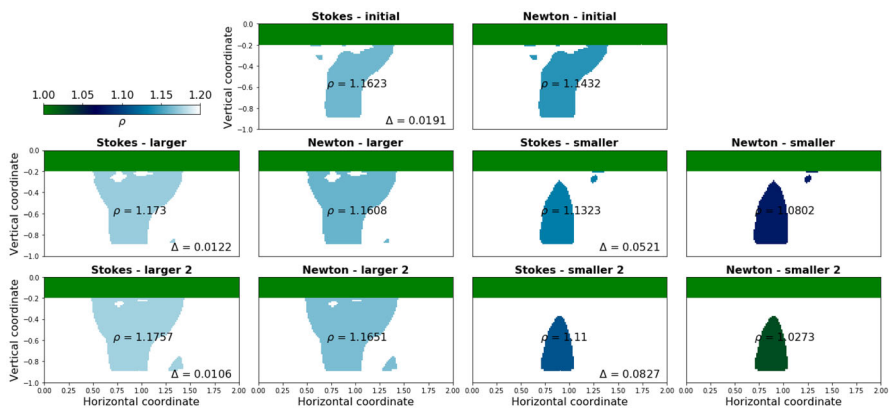
Fig. 18 Iterative density inversion in the mantle plume scenario, laid out and annotated as in Figs. 16 and 17

A competing algorithm will again start with the inversion for the wave speed and its binarization, but then normalize by their standard deviation and add the gradients of the density subject to gravitational potential and flow. This effective density gradient is then used for an inversion for the density structure minimizing the flow and the gravitational potential misfits. Simultaneously, the reference viscosity, subject to the flow misfit, is optimized. This offers added potential for imaging: one can choose a subdiscretization for the parameter space such that it is coarser than the solution space. Different levels of coarsening can reveal the robustness of structures.

The results of such an approach are presented in Fig. 19 for the pointwise gradient under the same discretization as the solution, whereas in Fig. 20 we used a fourfold
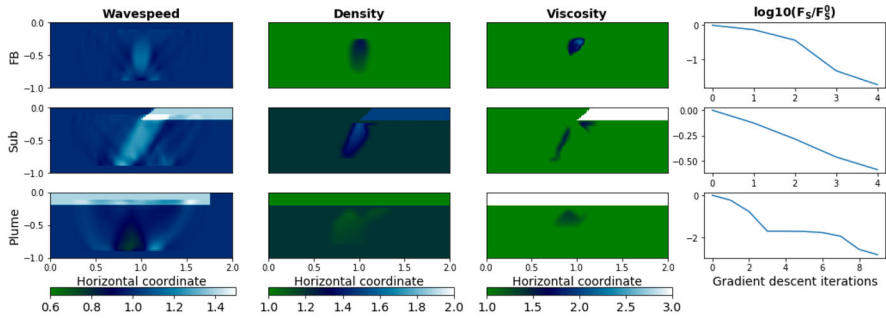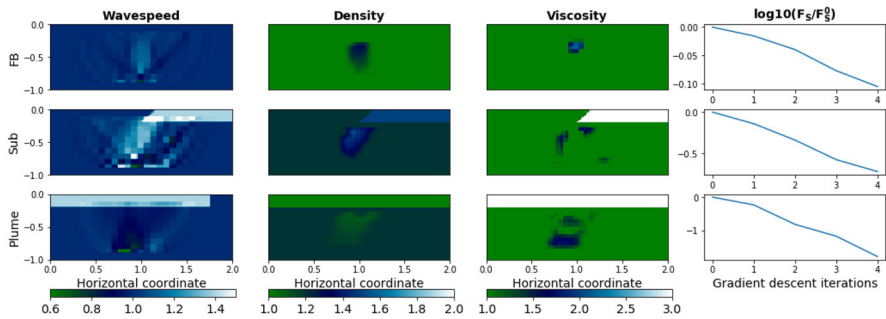
Fig. 19 Summary of the simultaneous inversion for the pointwise density and reference viscosity. The rows correspond to the scenarios falling block (FB), subduction zone (SZ) and mantle plume (MP). The columns show one of the resulting inverted material parameters, (i) the wave speed c subject to the acoustic wave equation, (ii) the density ρ masked by a contour line from the inverted wave speed, (iii) the reference viscosity eta_0 masked by a contour line from the inverted wave speed. The last column shows the convergence history for surface deformation from the Stokes equation as function of gradient-descent iterations

Fig. 20 Summary of the simultaneous inversion for a homogeneously $4\times$ coarsened density and reference viscosity. The rows correspond to the scenarios falling block (FB), subduction zone (SZ) and mantle plume (MP). The columns show one of the resulting inverted material parameters, (i) the wave speed c subject to the acoustic wave equation, (ii) the density ρ masked by a contour line from the inverted wave speed, (iii) the reference viscosity η_0 masked by a contour line from the inverted wave speed. The last column shows the convergence history for surface deformation from the Stokes equation as function of gradient-descent iterations

coarsening—rendering the gradient space four times smaller than the solution space, and in Fig. 21 we used an eightfold coarsening factor.

The inversion with factor 4 reveals the shapes of both block and mantle plume better than using an uncoarsened gradient. This algorithm, of course, has to be tested in terms of whether the observed improvements are consistently maintained in more complex settings.

The collinear gradient inversion described in Sect. 4.2.2, whereby parameter updates are applied where the gradients point in the same direction, can be applied in the iterative context of this section also, as shown in Fig. 22. In contrast to the previous iterative approaches the wave speed inversion is neglected here. As expected, the sole combination of the gravity and flow gradients for the densities and the viscosity does not result in a good image of the anomaly, but only recovers a very broad area
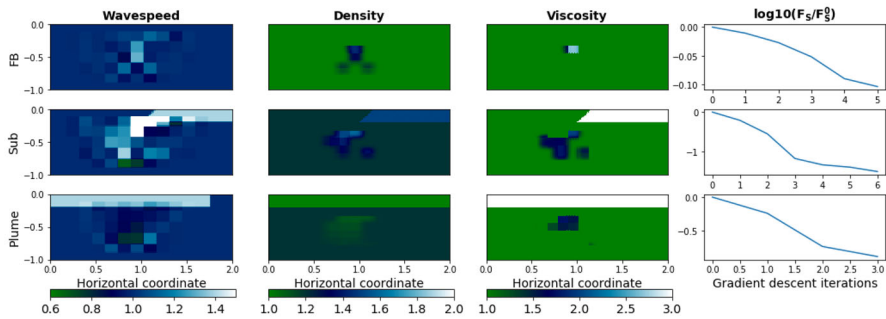
Fig. 21 Summary of the simultaneous inversion for a homogeneously $8\times$ coarsened density and reference viscosity. The rows correspond to one of the following scenarios, falling block (FB), subduction zone (SZ) and mantle plume (MP). The columns show one of the resulting inverted material parameter, (i) the wave speed c subject to the acoustic wave equation, (ii) the density ρ masked by a contour line from the inverted wave speed, (iii) the reference viscosity η_0 masked by a contour line from the inverted wave speed. The last column shows the convergence history for surface deformation from the Stokes equation as function of gradient-descent iterations
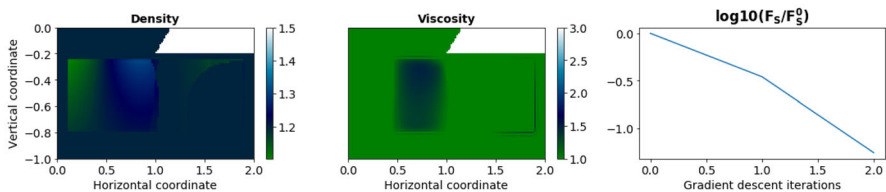

Fig. 22 Summary of the simultaneous inversion for the density and viscosity subject to the Stokes and Poisson equation, respectively, for the subduction zone scenario. Material parameters are only updated at points where all three gradients point in the same direction. The left panel shows the inverted density, the middle panel the inverted reference viscosity, and the right panel shows the convergence history for the surface deformation subject to the Stokes equations as function of gradient-descent iterations. For this inversion no acoustic wavefield information was considered

to the left of the synthetic subduction zone. This observation is however consistent with the kernels shown in Fig. 8, where the viscosity and the density show the highest sensitivity towards the top-left area of the synthetic anomaly.

4.3 Regularization

Until now in this paper, wave-equation information has been used to "precondition" subsequent inversions via contour-informed binarized masking of the gradients of the flow and gravitational constraints. Another approach is to use the initial wave speed inversion results as explicit regularization. Here, we employ a Tikhonov-type (Aster et al. 2005; Tikhonov and Arsenin 1977) regularization whereby the PDE-specific data-space cost functions F_β are penalized by the addition of a model-space quadratic norm, as

$$\tilde{F}_\beta = F_\beta + \frac{\vartheta}{2}\|\mathscr{R}(\mathbf{m})\|_2^2 + \frac{\varphi}{2}\|\mathbf{m}\|_2^2 , \tag{29}$$
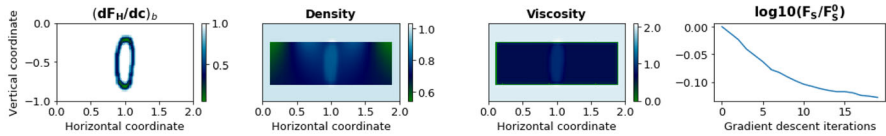
Fig. 23 Summary of the regularized simultaneous inversion for the pointwise density and viscosity subject to the Stokes and Poisson equations in the falling block scenario. Material parameters are only updated at points where all three gradients point in the same direction. The first panel shows the binarized gradient of the wave speed used in the regularization term. The second panel shows the density inversion result. The third panel shows the inverted reference viscosity. The fourth panel shows the convergence history for the surface deformation subject to the Stokes equations over the gradient-descent iterations

with ϑ and φ trade-off parameters, and $\mathscr{R}$ a diffusion-style (Scherzer and Weickert 2000; Villa et al. 2019) operator:

$$\mathscr{R} = \nabla \cdot \left[b \left(\left| \frac{\partial F_H}{\partial c} \right| \right) \nabla \right], \tag{30}$$

where $b \left(|\partial F_H / \partial c| \right)$ is the binarization of the pointwise derivative of the Hookean cost function in function of the wavespeed, as shown in the first panel of Fig. 23. Our aim is to bias the inversion into correlating updates with the wavespeed results. Results of such an inversion for the falling block case are shown in Fig. 23. In this initial experiment the updates for the density and the viscosity are not shedding much additional light on the shape or parameters of the anomaly, but we do keep this procedure in mind as a future research direction.

4.4 Shape recovery by parameterized inversion

In contrast to inverting for both geometry and material parameters of subsurface anomalies, geological situations arise where the parameter contrasts are known or can be assumed from prior knowledge, but the shape of the anomalies is unknown and remains of interest. For instance, rock type and seismic wavespeed perturbations in a subduction zone might be relatively well known, but the angle and shape of the zone might be unclear and the seismic station network very sparse.

An algorithm to tackle such an inverse problem, which has gained great popularity in engineering, is the deterministic level set method (Aghasi et al. 2011; Burger 2001; Dorn and Lesselier 2015). The "level set" is the common boundary $\delta \Omega_\phi$ of a smooth function within the domain Ω. The level set function ϕ vanishes on the boundary and decreases or increases away from it. In our application we take the signed distance of the normalized wavespeed result, $c(\mathbf{r})$, of the waveform inversion (as shown in, e.g., Fig. 14) as the initial level set function:

$$\phi(\mathbf{r}) = \text{sign} \left(\frac{c(\mathbf{r}) - \min[c(\mathbf{r})]}{\max[c(\mathbf{r})] - \min[c(\mathbf{r})]} \right). \tag{31}$$

Points where $\phi > 0$ are considered "subduction zone", whereas "background" mantle has $\phi \leq 0$, resulting in a indicator function χ for the material parameters, for instance

$$\rho(\mathbf{r}) = \chi_{\phi>0}\,\rho_{SZ} + \chi_{\phi<0}\,\rho_{BG}. \tag{32}$$

The objective is to change the level set (the shape of the anomaly), in order to improve the fit to the observations. The evolution is by a Hamilton–Jacobi equation with time τ an optimization time that proceeds according to the inversion iterations, as follows:

$$\frac{\partial \phi}{\partial \tau} + \mathbf{v} \cdot \nabla \phi = 0. \tag{33}$$

The aim is to obtain a boundary advection velocity $\mathbf{v}$ so as to result in a reduction of the target cost function. This is achieved by evaluating the gradient of the cost function with respect to the indicator function. Using the vector $\hat{\boldsymbol{\gamma}}$, the normal to the boundary $\delta\Omega_\phi$, and v_γ, the velocity component projected onto it,

$$\hat{\boldsymbol{\gamma}}(\delta\Omega_\phi, \tau) = \frac{\nabla \phi}{|\nabla \phi|} \quad \text{and} \quad v_\gamma = \mathbf{v} \cdot \hat{\boldsymbol{\gamma}}, \tag{34}$$

we simplify Eq. (33) to

$$\frac{\partial \phi}{\partial \tau} + v_\gamma\,|\nabla \phi| = 0, \tag{35}$$

to describe the evolution of the level set. The direction perpendicular to the boundary is taken to be the gradient of the cost function with respect to the material parameter. Where this gradient is negative, the size of the anomaly will be increased by advection of the level set away from the boundary, by an amount proportional to the magnitude of the gradient of the material parameter. In this work, the advection equation (35) is solved with a forward Euler scheme (Laurain 2018; Peng et al. 1999), care being taking to reinitialize the level set function every few iterations.

Figure 24 presents the result of the level-set inversion for the subduction zone. The initial guess for the shape of the anomaly, obtained from the wavespeed contour of an already quite successful seismic inversion, is already quite close to the truth, except it reaches too deep and extends too wide. Our aim is to improve the image by the addition of information from the vertical Stokes velocity at the surface. Three cases are considered: (i) where only the gradient of the velocity with respect to the density is used to advect the level set; (ii) where only the gradient of the velocity with respect to the viscosity is used as advection velocity, and (iii) where both gradients are used as advection velocity. Interestingly, the inversion using only the density gradient results in an anomaly that is thinner than the target. In contrast, the inversion using just the viscosity results in an anomaly that is too large. An inversion that uses both gradients rather nicely reproduces the upper parts of the anomaly, bending at the correct depth, but failing to reduce the vertical extent of the imaged structure.
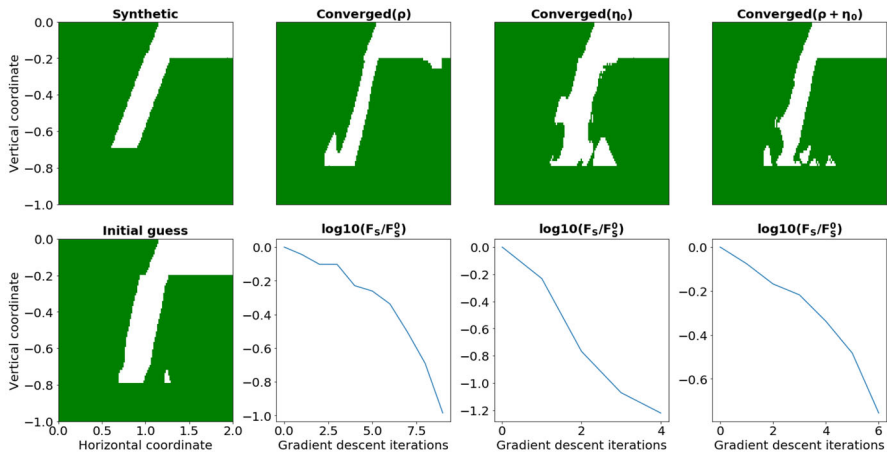
Fig. 24 Summary of the level set inversion with the vertical surface velocity as data under the subduction zone scenario. The panels in the upper row are, from left to right: the synthetic shape of the subduction zone to be reconstructed, represented as a binarized version of the level set function ϕ; the converged result using only the gradient of the density; the converged result using only the viscosity gradient; converged result using both the density and viscosity gradients. The panels in the row below show the initial guess for the geometry of the subduction zone based on a prior wave speed inversion, and the normalized cost functions over the gradient-descent iterations corresponding to the three cases in the panels above them

Using a seismic wave speed inversion result as the initial guess for a subsequent geometric level-set inversion driven by additional geophysical observables is a promising approach. While these few experiments have barely done it justice, we hope that our suggestions will find their way into future research programs.

5 Discussion

Sensitivity kernels and cost-function gradients highlight both strengths and weaknesses of the various geophysical observables that can be taken into consideration for the inversion for material parameters and the imaging of Earth structure.

The acoustic wave equation provides good initial sensitivity to the outline and geometry of wave speed anomalies that are *a priori* unknown. Hence, wave speed inversion of seismogram data, records of transient surface deformation, is recommended as the first step in a geophysical imaging study. Wave speed inversion results can be non-unique and marred by artifacts and spurious additional anomalies. Wave speeds themselves cannot be directly translated into rheological parameter values.

The gravitational potential observed at the Earth's surface, described by Poisson's equation driven by mass density perturbations, lacks direct resolving power or information on the geometry of the anomalies. Hence it is recommended to precondition or regularize inversions of potential-field data using predefined geometrical shapes inside of which a presumably slowly varying or even constant density becomes the unknown inversion parameter.

Tectonic surface deformation operates on very long timescales and hence is another essentially time-independent geophysical observable. The Stokes flow equations link surface velocity to density and viscosity anomalies inside the Earth. However, the gradients with respect to the density are only very diffusely illuminating the subsurface anomaly positions, and sometimes biasing their location towards the measurement surface. Viscosity gradients tend to be localized around the positions of viscosity jumps, thus also lacking direct and precise information on the location of unknown anomalies. Care must be taken to not simply update the viscosity at such positions in order to fit the observed data, which could lock the result into a local minimum of the cost function. Nevertheless, the flow equations are the sole set of the three PDEs considered in this paper that are also sensitive to the effective viscosity. Hence, minimally, they can act as a cross check on density inversions resulting from the gravitational potential.

Under optimal circumstances wave speed inversions provide multiple plausible initial guesses for the location and geometry of subsurface anomalies. The inversion for the density of this anomaly should then fit the gravity signal and at the same time the surface deformation. If the latter remains challenging the initial guess may not have been good enough. Otherwise, the flow equations can be used to, finally, deliver the effective viscosity of the anomaly.

6 Conclusion and outlook

Combining multiple geophysical observables and inverse methods into a coherent framework for the joint multiparameter imaging of subsurface structures has moved from theoretical considerations into the practical realm through the availability of flexible computational routines that allow for data assimilation and inversion under PDE constraints. The promise of improving the topology of individual cost functions by summing them into a combination, and of reducing the non-uniqueness plaguing single methods used individually, has motivated our work. We focused on three classes of geophysical fields: seismic wave propagation influenced by wave speed anomalies, the gravitational disturbing potential caused by density perturbation, and steady-state surface deformation controlled by density and viscosity structure, in three geological settings: a falling block, a subduction zone, and a mantle plume.

Our positive outlook is informed by (i) the comparison of low-dimensional grid-searches within the landscape of summed cost functions, showing those to be preferred over sampling individual cost functions separately; (ii) the calculation of the cost-function gradients and sensitivity kernels of each observable with respect to each design variable, unveiling that gradient combinations outperform sequential optimization; (iii) ideas of preconditioning and regularizing combined inversions, and for geometric level-set, shape-optimizing rather than pointwise approaches.

To reap the benefits of combined inversion approaches, researchers within geophysical subdomains should continue to make their data and models widely available, so that they can be used as initial models, additional constraints, preconditioners, or regularizers, across the disciplines. Our work has furthered the cause by laying out the technical pieces to efficiently calculate the gradients of the three geophysical observ-

ables described above, with respect to the design variables of density, wavespeed and viscosity, using an adjoint-state framework.

The simple examples and straightforward algorithms presented here do not point to a single "best-performing" coupled inversion procedure, but high-level languages such as FEniCS will continue to make it affordable for others to develop sandbox testing frameworks for combined inversions following the techniques and procedures described in this paper. All of our examples have been deterministic and gradient-based, stopping short of evaluating Hessian matrices to quantify uncertainties, but it is clear that the embedding of our procedures within more sophisticated statistical sampling approaches will open up fruitful new avenues of research.

We imagine a future in which additional observations such as magnetic anomaly measurements (Lewis and Simons 2012), electric resistivity probings (Robbins and Plattner 2018) and ground-penetrating radar (Domenzain et al. 2018; Plattner 2020), would form part and parcel of a holistic geophysical inversion scheme. Even within our own limited framework we imagine including additional complexity in the physical description of the Earth system, such as incorporating full elasticity and even anelastic viscous dampening of the seismic wavefield (Askan et al. 2007, 2010; Pan and Wang 2020; Tarantola 1986, 1988), including visco-elasto-plastic effects in the Stokes equation (Peltier 1985), and allowing for additional kernel evaluations, e.g., for internal friction, with additional coupling between Hooke's and Stoke's problems through the elastic shear modulus. However, our examples also illustrate that for a combined inversion approach to be truly proclaimed superior to a sequential one, it needs to be thoroughly tried and tested on real-world examples, which might impose on future studies additional technical developments.

References

Aghasi, A., Kilmer, M., Miller, E.L.: Parametric level set methods for inverse problems. SIAM J. Imag. Sci. **4**(2), 618–650 (2011). https://doi.org/10.1137/100800208

Akçelik, V., Biros, G., Ghattas, O.: Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation. In: Proceedings of Conference on Supercomputing, ACM/IEEE (2002). https://doi.org/10.1109/SC.2002.10002

Aki, K., Richards, P.G.: Quantitative Seismology, 1st edn. Freeman, San Francisco (1980)

Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., Wells, G.N.: The FEniCS Project version 1.5. Arch. Numer. Softw. **3**(100), 9–23 (2015)

Askan, A., Akcelik, V., Bielak, J., Ghattas, O.: Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures. Bull. Seismol. Soc. Am. **97**(6), 1990–2008 (2007). https://doi.org/10.1785/0120070079

Askan, A., Akcelik, V., Bielak, J., Ghattas, O.: Parameter sensitivity analysis of a nonlinear least-squares optimization-based anelastic full waveform inversion method. CR Mécanique **338**(7–8), 364–376 (2010). https://doi.org/10.1016/j.crme.2010.07.002

Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems. International Geophysics Series, vol. 90. Elsevier Academic Press, San Diego (2005)

Balay, S., Abhyankar, S., Adams, M., Brown, J., Brune, P., Buschelman, K., Dalcin, L., Dener, A., Eijkhout, V., Gropp, W., Karpeyev, D., Kaushik, D., Knepley, M., May, D., McInnes, L.C., Mills, R., Munson, T., Rupp, K., Sanan, P., Smith, B., Zampini, S., Zhang, H.: PETSc Users Manual. Revision 3.11. Technical report, Argonne National Laboratory, Chicago, IL (2019)

Berkel, P., Michel, V.: On mathematical aspects of a combined inversion of gravity and normal mode variations by a spline method. Math. Geosci. **42**(7), 795–816 (2010). https://doi.org/10.1007/s11004-010-9297-2

Blakely, R.J.: Potential Theory in Gravity and Magnetic Applications. Cambridge University Press, New York (1995)

Brown, J.R., Slawinski, M.A.: On Foundations of Seismology: Bringing Idealizations Down to Earth. World Scientific, Singapore (2017)

Bunge, H.-P., Richards, M.A., Baumgardner, J.R.: Mantle-circulation models with sequential data assimilation: inferring present-day mantle structure from plate-motion histories. Philos. Trans. R. Soc. London Ser. A **360**, 2545–2567 (2002). https://doi.org/10.1098/rsta.2002.1080

Bunge, H.-P., Hagelberg, C.R., Travis, B.J.: Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography. Geophys. J. Int. **152**(2), 280–301 (2003). https://doi.org/10.1046/j.1365-246X.2003.01823.x

Bunks, C., Saleck, F.M., Zaleski, S., Chavent, G.: Multiscale seismic waveform inversion. Geophysics **60**(5), 1457–1473 (1995). https://doi.org/10.1190/1.1443880

Burger, M.: A level set method for inverse problems. Inverse Probl. **17**(5), 1327 (2001). https://doi.org/10.1088/0266-5611/17/5/307

Chao, B.F.: On inversion for mass distribution from global (time-variable) gravity field. J. Geodyn. **39**, 223–230 (2005). https://doi.org/10.1016/j.jog.2004.11.001

Claerbout, J.F.: Earth Soundings Analysis: Processing Versus Inversion. Blackwell, Cambridge (1992)

Colli, L., Ghelichkhan, S., Bunge, H.-P., Oeser, J.: Retrodictions of Mid Paleogene mantle flow and dynamic topography in the Atlantic region from compressible high resolution adjoint mantle convection models: Sensitivity to deep mantle viscosity and tomographic input model. Gondwana Res. **53**, 252–272 (2018). https://doi.org/10.1016/j.gr.2017.04.027

Condie, K.C.: Mantle Plumes and Their Record in Earth History. Cambridge University Press, Cambridge (2001)

Conrad, C.P., Molnar, P.: The growth of Rayleigh–Taylor-type instabilities in the lithosphere for various rheological and density structures. Geophys. J. Int. **129**(1), 95–112 (1997). https://doi.org/10.1111/j.1365-246X.1997.tb00939.x

Conrad, C.P., Steinberger, B., Torsvik, T.H.: Stability of active mantle upwelling revealed by net characteristics of plate tectonics. Nature **498**(7455), 479–482 (2013). https://doi.org/10.1038/nature12203

Crestel, B., Stadler, G., Ghattas, O.: A comparative study of structural similarity and regularization for joint inverse problems governed by PDEs. Inverse Probl. **35**, 024003 (2018). https://doi.org/10.1088/1361-6420/aaf129

Dahlen, F.A., Tromp, J.: Theoretical Global Seismology. Princeton University Press, Princeton (1998)

de Hoop, M.V., Smith, H., Uhlmann, G., van der Hilst, R.D.: Seismic imaging with the generalized Radon transform: a curvelet transform perspective. Inverse Probl. **25**(2), 025005 (2009). https://doi.org/10.1088/0266-5611/25/2/025005

Domenzain, D., Bradford, J., Mead, J.: Joint inversion of GPR and ER data. In: SEG Technical Program Expanded Abstracts, Society of Exploration Geophysicists, Denver, CO, pp. 4763–4767 (2018). https://doi.org/10.1190/segam2018-2997794.1

Dorman, L.M., Lewis, B.T.R.: Experimental isostasy: 1. Theory of the determination of the Earth's isostatic response to a concentrated load. J. Geophys. Res. **75**(17), 3357–3365 (1970)

Dorn, O., Lesselier, D.: Level set methods for structural inversion and image reconstruction. In: Scherzer, O. (ed.) Handbook of Mathematical Methods in Imaging. Springer, New York (2015)

Elkins-Tanton, L.T.: Continental magmatism caused by lithospheric delamination. Geol. Soc. Am. Spec. Pap. **388**, 449–461 (2005)

Elkins-Tanton, L .T.: Continental magmatism, volatile recycling, and a heterogeneous mantle caused by lithospheric gravitational instabilities. J. Geophys. Res. **112**(B3), B03,405 (2007). https://doi.org/10.1029/2005JB004072

Fichtner, A., Simutė, S.: Hamiltonian Monte Carlo inversion of seismic sources in complex media. J. Geophys. Res. **123**(4), 2984–2999 (2018). https://doi.org/10.1002/2017JB015249

Fichtner, A., Trampert, J.: Hessian kernels of seismic data functionals based upon adjoint techniques. Geophys. J. Int. **185**(2), 775–798 (2011). https://doi.org/10.1111/j.1365-246X.2011.04966.x

Fichtner, A., Bunge, H.-P., Igel, H.: The adjoint method in seismology—II Applications: traveltimes and sensitivity functionals. Phys. Earth Planet. Inter. **157**(1–2), 105–123 (2006a). https://doi.org/10.1016/j.pepi.2006.03.018

Fichtner, A., Bunge, H.-P., Igel, H.: The adjoint method in seismology—I. Theory. Phys. Earth Planet. Inter. **157**(1–2), 86–104 (2006b). https://doi.org/10.1016/j.pepi.2006.03.016

Fischer, D., Michel, V.: Sparse regularization of inverse gravimetry–case study: spatial and temporal mass variations in South America. Inv. Probl. **28**, 065012 (2012). https://doi.org/10.1088/0266-5611/28/6/065012

Forte, A.M., Mitrovica, J.X.: Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data. Nature **410**(6832), 1049–1056 (2001). https://doi.org/10.1038/35074000

Freeden, W.: Chapter 1: Geomathematics: its role, its aim, and its potential. In: Freeden, W., Nashed, M.Z.Z., Sonar, T. (eds.) Handbook of Geomathematics, pp. 3–78. Springer, Berlin (2015). https://doi.org/10.1007/978-3-642-01546-5_1

Freeden, W., Nashed, M.Z.: Inverse gravimetry: background material and multiscale mollifier approaches. Int. J. Geomath. **9**(2), 199–264 (2018). https://doi.org/10.1007/s13137-018-0103-5

Gauthier, O., Virieux, J., Tarantola, A.: Two-dimensional nonlinear inversion of seismic waveforms: numerical results. Geophysis **51**(7), 1387–1403 (1986)

Geng, Y., Innanen, K., Pan, W.: Subspace method for multi-parameter FWI. Commun. Math. Phys. **28**, 228–248 (2020). https://doi.org/10.4208/cicp.OA-2018-0087

Gerya, T.: Introduction to Numerical Geodynamic Modelling. Cambridge University Press, Cambridge (2019)

Ghelichkhan, S., Bunge, H.-P.: The compressible adjoint equations in geodynamics: derivation and numerical assessment. Int. J. Geomath. **7**(1), 1–30 (2016). https://doi.org/10.1007/s13137-016-0080-5

Glatzmaier, G.A.: Introduction to Modeling Convection in Planets and Stars: Magnetic Field, Density Stratification, Rotation. Princeton University Press, Princeton (2014)

Gouveia, W.P., Scales, J.A.: Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. J. Geophys. Res. **103**(B2), 2759–2779 (1998). https://doi.org/10.1029/97JB02933

Harig, C., Zhong, S., Simons, F .J.: Constraints on upper-mantle viscosity inferred from the flow-induced pressure gradient across a continental keel. Geochem. Geophys. Geosyst. **11**(6), Q06,004 (2010). https://doi.org/10.1029/2010GC003038

Hofmann-Wellenhof, B., Moritz, H.: Physical Geodesy, 2nd edn. Springer, New York (2006)

Horbach, A., Bunge, H.-P., Oeser, J.: The adjoint method in geodynamics: derivation from a general operator formulation and application to the initial condition problem in a high resolution mantle circulation model. Int. J. Geomath. **5**(2), 163–194 (2014). https://doi.org/10.1007/s13137-014-0061-5

Kellogg, O.D.: Foundations of Potential Theory. Springer, New York (1967)

Kennett, B.L.N., Bunge, H.-P.: Geophysical Continua. Cambridge University Press, Cambridge (2008)

Laurain, A.: A level set-based structural optimization code using FEniCS. Struct. Multidiscip. Optim. **58**(3), 1311–1334 (2018). https://doi.org/10.1007/s00158-018-1950-2

Lewis, K .W., Simons, F .J.: Local spectral variability and the origin of the Martian crustal magnetic field. Geophys. Res. Lett. **39**, L18,201 (2012). https://doi.org/10.1029/2012GL052708

Lithgow-Bertelloni, C., Richards, M.A.: The dynamics of Cenozoic and Mesozoic plate motions. Rev. Geophys. **36**, 27–78 (1998)

Liu, Q., Tromp, J.: Finite-frequency sensitivity kernels based upon adjoint methods. Bull. Seismol. Soc. Am. **96**(6), 2383–2397 (2006). https://doi.org/10.1785/0120060041

Liu, Q., Tromp, J.: Finite-frequency sensitivity kernels for global seismic wave equation based upon adjoint methods. Geophys. J. Int. **174**, 265–286 (2009). https://doi.org/10.1111/j.1365-246X.2008.03798.x

Logg, A., Andre-Mardal, K., Wells, G.N.: Automated solution of differential equations by the finite element method: The FEniCS book. Lecture Notes in Computational Science and Engineering, vol. 84. Springer, Berlin (2012)

Ma, Y., Hale, D.: Quasi-Newton full-waveform inversion with a projected Hessian matrix. Geophysics **77**(5), R207–R216 (2012). https://doi.org/10.1190/GEO2011-0519.1

Malevsky, A.V., Yuen, D.A.: Strongly chaotic non-Newtonian mantle convection. Geophys. Astrophys Fluid Dyn. **65**(1–4), 149–171 (1992)

Malvern, L.E.: Introduction to the Mechanics of a Continuous Medium. Prentice-Hall, Englewood Cliffs (1969)

Martin, D., Nokes, R.: Crystal settling in a vigorously convecting magma chamber. Nature **332**(6164), 534–536 (1988). https://doi.org/10.1038/332534a0

Martin, D., Nokes, R.: A fluid-dynamical study of crystal settling in convecting magmas. J. Petrol. **30**(6), 1471–1500 (1989). https://doi.org/10.1093/petrology/30.6.1471

Martin, J., Wilcox, L.C., Burstedde, C., Ghattas, O.: A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. SIAM J. Sci. Comput. **34**(3), A1460–A1487 (2012). https://doi.org/10.1137/110845598

Mead, J.L., Ford, J.F.: Joint inversion of compact operators. J. Inverse Ill-posed Probl. **28**(1), 105–118 (2020). https://doi.org/10.1515/jiip-2019-0068

Melosh, H.J., Raefsky, A.: The dynamical origin of subduction zone topography. Geophys. J. Int. **60**(3), 333–354 (1980). https://doi.org/10.1111/j.1365-246X.1980.tb04812.x

Michel, V.: Regularized wavelet-based multiresolution recovery of the harmonic mass density distribution from data of the Earth's gravitational field at satellite height. Inverse Probl. **21**, 997–1025 (2005). https://doi.org/10.1088/0266-5611/21/3/013

Michel, V.: Chapter 32: Tomography: problems and multiscale solutions. In: Freeden, W., Nashed, M.Z., Sonar, T. (eds.) Handbook of Geomathematics, 2nd edn, pp. 2087–2119. Springer, Berlin (2015). https://doi.org/10.1007/978-3-642-01546-5_32

Michel, V., Fokas, A.S.: A unified approach to various techniques for the non-uniqueness of the inverse gravimetric problem and wavelet-based methods. Inverse Probl. **24**, 045019 (2008). https://doi.org/10.1088/0266-5611/24/4/045019

Michel, V., Orzlowski, S.: On the nullspace of a class of Fredholm integral equations of the first kind. J. Inverse Ill-posed Probl. **24**(6), 1–24 (2015). https://doi.org/10.1515/jiip-2015-0026

Michel, V., Simons, F.J.: A general approach to regularizing inverse problems with regional data using Slepian wavelets. Inverse Probl. **33**(12), 12,501 (2017). https://doi.org/10.1088/1361-6420/aa9909

Mora, P.: Elastic wave-field inversion of reflection and transmission data. Geophysics **53**(6), 750–759 (1988)

Mora, P.: Inversion = migration + tomography. Geophysics **54**(12), 1575–1586 (1989)

Morgan, W .J.: Gravity anomalies and convection currents. 1. A sphere and cylinder sinking beneath surface of a viscous fluid. J. Geophys. Res. **70**(24), 6175–6185 (1965). https://doi.org/10.1029/JZ070i024p06,175

Morgan, W.J.: Convection plumes in the lower mantle. Nature **230**(5288), 42–43 (1971). https://doi.org/10.1038/230,042a0

Nolet, G.: A Breviary for Seismic Tomography. Cambridge University Press, Cambridge (2008)

Nolet, G.: Transmission tomography in seismology. In: Freeden, W., Nashed, M.Z., Sonar, T. (eds.) Handbook of Geomathematics, 2nd edn, pp. 1887–1904. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-642-54551-1_58

Oldenburg, D.W.: The inversion and interpretation of gravity anomalies. Geophysics **39**(4), 526–536 (1974). https://doi.org/10.1190/1.1440,444

Pan, W., Wang, Y.: On the influence of different misfit functions for attenuation estimation in viscoelastic full-waveform inversion: synthetic study. Geophys. J. Int. **221**(2), 1292–1319 (2020). https://doi.org/10.1093/gji/ggaa089

Parker, R.L.: The rapid calculation of potential anomalies. Geophys. J. R. Astron. Soc. **31**(4), 447–455 (1973). https://doi.org/10.1111/j.1365-246X.1973.tb06513.x

Parmentier, E.M., Turcotte, D.L., Torrance, K.E.: Studies of finite amplitude non-Newtonian thermal convection with application to convection in the Earth's mantle. J. Geophys. Res. **81**(11), 1839–1846 (1976). https://doi.org/10.1029/JB081i011p01839

Peltier, W.R.: Mantle convection and viscoelasticity. Annu. Rev. Fluid Mech. **17**(1), 561–608 (1985). https://doi.org/10.1146/annurev.fl.17.010185003021

Peng, D., Merriman, B., Osher, S., Zhao, H., Kang, M.: A PDE-based fast local level set method. J. Comput. Phys. **155**(2), 410–438 (1999). https://doi.org/10.1006/jcph.1999.6345

Petra, N., Zhu, H., Stadler, G., Hughes, T.J.R., Ghattas, O.: An inexact Gauss–Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model. J. Glaciol. **58**(211), 889–903 (2012). https://doi.org/10.3189/2012JoG11J182

Petra, N., Martin, J., Stadler, G., Ghattas, O.: A computational framework for infinite-dimensional Bayesian inverse problems, Part II: stochastic Newton MCMC with application to ice sheet flow inverse problems. SIAM J. Sci. Comput. **36**(4), A1525–A1555 (2014). https://doi.org/10.1137/130934,805

Plattner, A.M.: GPRPy: open-source ground-penetrating radar processing and visualization software. Lead. Edge **39**(5), 332–337 (2020). https://doi.org/10.1190/tle39050,332.1

Plessix, R.-E.: A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophys. J. Int. **167**(2), 495–503 (2006). https://doi.org/10.1111/j.1365-246X.2006.02,978.x

Ranalli, G.: Rheology of the Earth, 2nd edn. Chapman and Hall, London (1995)

Ratnaswamy, V., Stadler, G., Gurnis, M.: Adjoint-based estimation of plate coupling in a non-linear mantle flow model: theory and examples. Geophys. J. Int. **202**(2), 768–786 (2015). https://doi.org/10.1093/gji/ggv166

Reuber, G., Holbach, L., Popov, A., Hanke, M., Kaus, B.: Inferring rheology and geometry of subsurface structures by adjoint-based inversion of principal stress directions. Geophys. J. Int. **223**(2), 851–861 (2020). https://doi.org/10.1093/gji/ggaa344

Reuber, G.S., Kaus, B.J.P., Popov, A.A., Baumann, T.S.: Unraveling the physics of the Yellowstone magmatic system using geodynamic simulations. Front. Earth Sci. **6**, 117 (2018). https://doi.org/10.3389/feart.2018.00117

Robbins, A.R., Plattner, A.: Offset-electrode profile acquisition strategy for electrical resistivity tomography. J. Appl. Geoph. **151**, 66–72 (2018). https://doi.org/10.1016/j.jappgeo.2018.01.027

Scherzer, O., Weickert, J.: Relations between regularization and diffusion filtering. J. Math. Imaging Vis. **12**(1), 43–63 (2000). https://doi.org/10.1023/A:1008344608808

Schubert, G., Turcotte, D.L., Olson, P.: Mantle Convection in the Earth and Planets. Cambridge University Press, Cambridge (2001)

Schuster, G.T.: Seismic Inversion. Society of Exploration Geophysicists, Tulsa (2017)

Sheriff, R.E., Geldart, L.P.: Exploration Seismology. Cambridge University Press, Cambridge (1995)

Sleep, N.H.: Stress and flow beneath island arcs. Geophys. J. Int. **42**(3), 827–857 (1975). https://doi.org/10.1111/j.1365-246X.1975.tb06454.x

Sleep, N.H.: Hotspots and mantle plumes: some phenomenology. J. Geophys. Res. **95**(B5), 6715–6736 (1990). https://doi.org/10.1029/JB095iB05p06,715

Stefanov, P., Uhlmann, G., Vasy, A., Zhou, H.: Travel time tomography. Acta Math. Sin. **35**(6), 1085–1114 (2019). https://doi.org/10.1007/s10,114-019-8338-0

Stern, R.J.: Subduction zones. Rev. Geophys. (2002). https://doi.org/10.1029/2001RG000108

Symes, W .W.: The seismic reflection inverse problem. Inverse Probl. **25**(12), 123008 (2009). https://doi.org/10.1088/0266-5611/25/12/123008

Tarantola, A.: Linearized inversion of seismic reflection data. Geophys. Prospect. **34**(2), 998–1015 (1984a). https://doi.org/10.1111/j.1365-2478.1984.tb00751.x

Tarantola, A.: Inversion of seismic reflection data in the acoustic approximation. Geophysics **49**(8), 1259–1266 (1984b). https://doi.org/10.1190/1.1441754

Tarantola, A.: A strategy for nonlinear elastic inversion of seismic reflection data. Geophysics **51**(10), 1893–1903 (1986)

Tarantola, A.: Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. Pure Appl. Geophys. **128**(1–2), 365–399 (1988)

Tikhonov, A .N., Arsenin, V .Y.: Solutions of Ill-Posed Problems. V. H. Winston & Sons, Washington (1977)

Tröltzsch, F.: Optimal Control of Partial Differential Equations: Theory, Methods and Applications. Graduate Studies in Mathematics, vol. 112. American Mathematical Society, Providence (2010)

Tromp, J.: Seismic wavefield imaging of Earth's interior across scales. Nat. Rev. Earth Environ. **1**, 40–53 (2020). https://doi.org/10.1038/s43017-019-0003-8

Tromp, J., Tape, C., Liu, Q.: Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. Geophys. J. Int. **160**(1), 195–216 (2005). https://doi.org/10.1111/j.1365-246X.2004.02453.x

van den Berg, A.P., van Keken, P., Yuen, D.A.: The effects of a composite non-Newtonian and Newtonian rheology on mantle convection. Geophys. J. Int. **115**(1), 62–78 (1993). https://doi.org/10.1111/j.1365-246X.1993.tb05588.x

Villa, U., Petra, N., Ghattas, O.: hIPPYlib: An extensible software framework for large-scale inverse problems gioverned by PDEs: Part I: Deterministic inversion and linearized Bayesian inference. arXiv:1909.03948 (2019)

Virieux, J., Operto, S.: An overview of full-waveform inversion in exploration geophysics. Geophysics **74**(6), WCC1–WCC26 (2009). https://doi.org/10.1190/1.3238367

Wilson, J.T.: Mantle plumes and plate motions. Tectonophysics **19**(2), 149–164 (1973). https://doi.org/10.1016/0040-1951(73)90037-1

Woodward, M.J.: Wave-equation tomography. Geophysics **57**(1), 15–26 (1992). https://doi.org/10.1190/1.1443179

Xu, P.: Determination of surface gravity anomalies using gradiometric observables. Geophys. J. Int. **110**, 321–332 (1992a)

Xu, P.: The value of minimum norm estimation of geopotential fields. Geophys. J. Int. **111**, 170–178 (1992b)

Yilmaz, Ö.: Seismic Data Analysis, vol. 1-2. Society of Exploration Geophysicists, Tulsa (2001)

Yuan, Y .O., Simons, F.J.: Multiscale adjoint waveform-difference tomography using wavelets. Geophysics **79**(3), WA79–WA95 (2014). https://doi.org/10.1190/GEO2013-0383.1

Zienkiewicz, O.C.: The Finite Element Method, 3rd edn. McGraw-Hill, London (1977)