Shape reconstruction in seismic full waveform inversion using a level set approach and time reversal

DOI:
10.1016/j.jcp.2020.110059

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Journal of Computational Physics

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
A level set based shape evolution approach is presented for the inversion of elastic waveform data with a special emphasis on the application to seismic full waveform inversion. A conservative representation of the linear elastic equation of motion in 2D is formulated as a symmetric hyperbolic system. The minimization of a suitably chosen least-squares data misfit functional is then performed by a shape evolution approach. This evolution is driven by Kaczmarz type gradient-based descent directions which are practically obtained by solving a time-reversed form of the same conservative elastic system. A level set method is employed for the computational description of the evolving shapes. Different regularization schemes are tested and compared for stabilizing this shape evolution, combined with an additional integrated optimization loop for simultaneously estimating different internal elastic parameter values. Numerical experiments in 2D are presented which demonstrate the performance of this novel shape reconstruction scheme in the application of salt dome reconstruction in seismic full waveform inversion.

© 2020 Elsevier Inc. All rights reserved.
times as part of a nonlinear elastic optimization scheme. Even with modern computer architectures, this can be a daunting task such that often simplified models are used, either for the forward simulation or for the reservoir model itself. For example, in most practical scenarios nowadays, a simplified 3D acoustic approximation is employed for addressing the underlying parameter estimation problem. This approach speeds up the inversion process, but comes with several disadvantages. First, only one of the many seismic (visco-)elastic parameters is actually reconstructed in the acoustic model, which means that important details of the actual reservoir model are missed out in the inversion. Second, the wave propagation process itself is modelled incorrectly, since mode conversion and any shear wave propagation are not captured correctly by this simplified acoustic model. Therefore, acoustic inversion of seismic FWI data is very limited and likely to yield incorrect and suboptimal results. Therefore, recent research has shifted towards the use of full elastic wave models for data inversion \cite{12, 13, 15, 16, 17}.

Certainly, using a more realistic full elastic model also comes with the difficulty that in the most accurate model a significant number of independent elastic parameters are required as input, which need to be estimated simultaneously as space-dependent functions from the given elastic data set. This distributed parameter estimation task is extremely difficult, and practical approaches are still under development. Amongst those difficulties to be addressed are the unavoidable cross-talk between those many independent parameters and the ill-posedness of the underlying inverse problem, including slow convergence and many local minima. Ideally, any available prior information of the reservoir model should be taken into account when addressing this difficult problem in order to stabilize the inversion and reduce the uncertainty of the obtained results. Such prior information can practically be incorporated into reconstruction algorithms in form of tailor-made regularization techniques.

For example, many of the above-mentioned applications aim to identify or track regions of anomalous elastic medium parameters embedded in some inhomogeneous background. Often, there are sharp discontinuities between those anomalous regions and the background, and the internal profile inside those anomalies is either of no great importance, or it does not vary significantly compared to the contrast to the background. In those applications, a shape-based regularization approach has been proposed in the literature for the representation of those anomalous regions. In particular, in the application of imaging ground structure with solid salt domes embedded, different level set based shape reconstruction approaches have been proposed recently in \cite{18, 19, 20, 21, 22, 23, 24, 25}.

We will as well address this important problem by using a level set representation of shapes. In the listed literature so far, a detailed analysis of the level set approach when using a full linear elastic wave velocity model is still missing. In the work presented here, we will follow the work initiated in \cite{25} and demonstrate how an efficient seismic FWI approach can be designed for situations in which the subsurface contains such solid bodies of salt, using a linear elastic system model. We will provide a proof-of-concept study by using a 2D conservative formulation of the linear elastic wave equation in an isotropic setting. This task can still be solved on modern standard desktop or laptop computers, but contains already most of the essential features of more realistic scenarios which would require large computer clusters and parallelized computer codes for efficient and correct data processing. With only slight modifications our method should also be applicable to other applications, such as for example the monitoring of CO2 sequestration sites.

In our elastic approach, three different elastic parameters are to be reconstructed from the given data. Using a shape-based method, we assume that all three parameter profiles, which are space dependent functions in general, share the same discontinuity set which coincides with the boundaries of the salt domes. Inside those salt bodies, each of the elastic parameters can take a different and unknown value, but all three of them are assumed here space independent inside those structures. Outside the salt domes, in general all three parameters can be fully space-dependent functions. The reconstruction of those external parameter profiles can be integrated in the inversion algorithm proposed here, but we will restrict ourselves in this proof-of-concept study to the situation where all three elastic parameter functions are given outside the embedded unknown shapes. In this setting, the seismic FWI problem reduces to a shape reconstruction problem which can be solved in a very stable and efficient way by employing a so-called level-set technique \cite{26, 27, 28}.

One benefit of employing such a shape-based approach is that the dimension of the space of unknown parameters is reduced significantly since instead of a complete internal profile of the salt domes, as in standard approaches, only the overall shapes of these bodies need to be determined from the data. Practically, instead of volumetric parameter derivatives inside this salt body, only shape derivatives or equivalent restricted narrow band volumetric derivatives close to the boundaries of the salt bodies need to be determined and applied. This not only stabilizes the overall inversion, but also helps to concentrate updates on local areas of interest.

Also the level set approach requires us to estimate descent directions. As usual in such large scale problems, an adjoint technique is employed in order to calculate those descent directions efficiently. However, in contrast to
many traditional approaches, we employ here a conservative formulation of elastic wave propagation for the gradient calculation whose main operator is mathematically skew-self-adjoint. This means that the application of the formal adjoint in time domain amounts to a simple and physically very intuitive time-reversal process. This comes with the additional advantage that the same elastic wave propagation model can be employed for the adjoint (and therefore gradient) calculation as it is anyway required for simulating seismic waves in the domain of interest for calculating the predicted data mismatch. The gradient calculation then follows a modified form involving time derivatives of fields instead of space derivatives as in many more traditional approaches. In this paper we show that this modified approach results in a stable and efficient optimization workflow which converges to very useful results.

Summarizing, the novelty of our approach is the combination of a (when extended to 3D) very realistic elastic wave propagation model with a shape-based regularization approach and a conservative time-reversal scheme for efficiently calculating descent directions of a specially designed data misfit cost functional. Even though our theoretical derivation and computational implementation in this paper are both focusing on a 2D setting due to restrictions to available computing resources, we are very confident that the overall scheme can easily be extended to tackling more complex situations in 3D. We anticipate that also attenuation and anisotropic effects of elastic wave propagation can be included in this scheme, with the necessary modifications properly applied to the algorithm. These features, however, are not yet included in the derivation presented here.

The remainder of this paper is organized as follows. In Section 2, the forward model, including the conservative elastic system formulation, is introduced. In Section 3 this elastic system is then reformulated as a symmetric hyperbolic system, and the corresponding inverse problem is formulated and addressed using the time-reversal concept for adjoint field propagation. The level set based shape reconstruction approach is described in details in Section 4, where also some more technical details of the proposed reconstruction algorithm are explained. In Section 5, we present the predicted data mismatch. The gradient calculation then follows a modified form involving time derivatives of fields instead of space derivatives as in many more traditional approaches. In this paper we show that this modified approach results in a stable and efficient optimization workflow which converges to very useful results.

Summarizing, the novelty of our approach is the combination of a (when extended to 3D) very realistic elastic wave propagation model with a shape-based regularization approach and a conservative time-reversal scheme for efficiently calculating descent directions of a specially designed data misfit cost functional. Even though our theoretical derivation and computational implementation in this paper are both focusing on a 2D setting due to restrictions to available computing resources, we are very confident that the overall scheme can easily be extended to tackling more complex situations in 3D. We anticipate that also attenuation and anisotropic effects of elastic wave propagation can be included in this scheme, with the necessary modifications properly applied to the algorithm. These features, however, are not yet included in the derivation presented here.

The remainder of this paper is organized as follows. In Section 2, the forward model, including the conservative elastic system formulation, is introduced. In Section 3 this elastic system is then reformulated as a symmetric hyperbolic system, and the corresponding inverse problem is formulated and addressed using the time-reversal concept for adjoint field propagation. The level set based shape reconstruction approach is described in details in Section 4, where also some more technical details of the proposed reconstruction algorithm are explained. In Section 5, we present first the numerical setup used in our numerical experiments. Then, the results of the numerical experiments are discussed which demonstrate the general behaviour of the proposed novel scheme on synthetic test cases addressing, in particular, the problem of reconstructing the shapes of salt domes from FWI data. Finally, Section 6 provides a brief summary, conclusions and some suggestions for future research.

2. The Forward Problem

The equation of motion for seismic elastic wave propagation can be written (using the Einstein summation convention which implicitly sums up over repeated indices on the right) in the form

\[
\rho \frac{d^2 u_i}{dt^2} = \frac{\partial \sigma_{ij}}{\partial x_j} + F_i, \quad i = 1, 2, 3
\]  

where \( t \) indicates time, \( x_j \) are the spatial variables for \( j = 1, 2, 3 \), \( F_i \) indicate the different components of the body force per unit volume, \( \rho \) is the mass density, \( \sigma_{ij} \) are the components of the second-order stress tensor \( \sigma \) and \( u_i \) are the cartesian components of the object displacement \( u \). In an isotropic medium, neglecting any visco-elastic effects, the stress tensor takes the form

\[
\sigma_{ij} = \lambda(x) \frac{\partial u_k}{\partial x_k} \delta_{ij} + \mu(x) \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right),
\]

where \( \lambda \) and \( \mu \) are the Lamé parameters. Here we have used the standard notation \( \delta_{ij} = 1 \) if \( i = j \) and \( \delta_{ij} = 0 \) if \( i \neq j \). When defining the tensor \( \epsilon \) by

\[
\epsilon_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right),
\]

(3)

obtains the form

\[
\rho \frac{d^2 u_i}{dt^2} = \frac{\partial}{\partial x_i} \left( \lambda(x) \nabla \cdot u \right) + \frac{\partial}{\partial x_j} \mu(x) \epsilon_{ij} + F_i, \quad i = 1, 2, 3.
\]  

(4)

Let us introduce, for the further analysis,

\[
v = \frac{\partial u}{\partial t},
\]

(5)

\[
p = \lambda \nabla \cdot u,
\]

(6)
where \( v \) indicates velocity and \( p \) essentially describes pressure. Then we can formulate two different descriptions for
the seismic elastic wave equations. First, from (1) and (2), we obtain the system in \((v, \sigma)\) formulation

\[
\begin{align*}
\frac{\partial v_i}{\partial t} &= \frac{1}{\rho} \left( \frac{\partial \sigma_{ij}}{\partial x_j} + F_i \right), \\
\frac{\partial \sigma_{ij}}{\partial t} &= \lambda(x) \frac{\partial v_i}{\partial x_k} \delta_{ij} + \mu(x) \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right),
\end{align*}
\]

(7)

(8)

where all medium parameters appear on the right-hand side of the system, mixed with spatial derivatives. Second, we
can derive the alternative system in \((v, \epsilon, p)\) formulation

\[
\begin{align*}
\rho \frac{\partial v_i}{\partial t} &= \frac{\partial p}{\partial x_i} + \frac{\partial \epsilon_{ij}}{\partial x_j} + F_i, \\
\frac{1}{\mu} \frac{\partial \epsilon_{ij}}{\partial t} &= \frac{\partial v_i}{\partial x_j} \frac{\partial v_j}{\partial x_i}, \\
\frac{1}{\lambda} \frac{\partial p}{\partial t} &= \nabla \cdot v
\end{align*}
\]

(9)

(10)

(11)

where all medium parameters appear on the left-hand side of the system in front of the time derivatives. Both formul-
ations can be used for discretizing the forward problem and for deriving efficient algorithms for the inverse problem,
in principle. Historically, the former system (7), (8) has been more popular for implementing elastic wave simulators
and accordingly for deriving optimization schemes for the medium parameters. Recently it has been indicated in
the literature that the alternative system (9)-(11) (or similar ‘conservative’ formulations) lead to very intuitive and
computationally convenient formulations when addressing the inverse problem, see for example [31, 32, 33, 6]. In the
present paper, we will follow this second route rigorously and will demonstrate that highly efficient solution methods
can be obtained this way.

Let us concentrate in the following on the two-dimensional case, where it is convenient to identify \((i, j) = (x, y)\) in
the notation. Then, the \((v, \sigma)\) system (7), (8) can be written explicitly (from now on without resorting to the Einstein
summation convention) in the form

\[
\begin{align*}
\frac{\partial v_x}{\partial t} &= \frac{1}{\rho} \left( \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + F_x \right), \\
\frac{\partial v_y}{\partial t} &= \frac{1}{\rho} \left( \frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + F_y \right), \\
\frac{\partial \sigma_{xx}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_x}{\partial x} + \lambda \frac{\partial v_y}{\partial y}, \\
\frac{\partial \sigma_{yy}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_y}{\partial x} + \lambda \frac{\partial v_x}{\partial y}, \\
\frac{\partial \sigma_{xy}}{\partial t} &= \mu \left( \frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right).
\end{align*}
\]

(12)

(13)

(14)

(15)

(16)

As mentioned, this system (12)-(16) is the starting point for many numerical implementations for modelling elastic
waves in 2D (and analogously in 3D) and for many classical inversion routines, see for example [14, 13, 34]. On the
other hand, the \((v, \epsilon, p)\) system (7)-(9) gets the form

\[
\begin{align*}
\rho \frac{\partial v_x}{\partial t} &= \frac{\partial \epsilon_{xx}}{\partial x} + \frac{\partial \epsilon_{xy}}{\partial y} + \frac{\partial p}{\partial x} + F_x, \\
\rho \frac{\partial v_y}{\partial t} &= \frac{\partial \epsilon_{xy}}{\partial x} + \frac{\partial \epsilon_{yy}}{\partial y} + \frac{\partial p}{\partial y} + F_y, \\
\frac{1}{2\mu} \frac{\partial \epsilon_{xx}}{\partial t} &= \frac{\partial v_x}{\partial x}, \\
\frac{1}{2\mu} \frac{\partial \epsilon_{yy}}{\partial t} &= \frac{\partial v_y}{\partial y},
\end{align*}
\]

(17)

(18)

(19)

(20)
where all medium parameters appear in front of the time derivatives. When solving a time-domain inverse problem as considered here, based on efficient adjoint techniques, we prefer the second formulation for various reasons. First, this ‘conservative formulation’ allows interpreting the corresponding adjoint system as a realization of a time-reversal process, which is physically intuitive. Second, usually, the same forward modelling code can be used for implementing the time-reversed system as for the forward problem, removing the need for having two different solvers for the direct and adjoint problem. Notice also that the formulations for gradient calculations based on (17)-(22) significantly differ from those derived from the system (12)-(16) because the medium parameters appear at different places in the system, and therefore interact with different terms of the system. This will have an impact on the general behaviour of the optimization process, which is worthwhile to be mentioned here.

We want to advocate one more motivation to use the system (17)-(22) instead of (12)-(16) in the application of seismic FWI. Actually, various alternative ‘conservative formulations’ have been proposed in the literature for the seismic FWI problem [31, 33, 6]. Notice that we will use in this paper the somehow ‘non-standard’ conservative formulation (17)-(22), following the general ideas outlined in [35] and further studied in a time-reversal context in [32]. One of the reasons is that in the resulting system (17)-(22) the pressure fields are explicitly included and calculated. In many practical applications of off-shore seismic FWI, seismic signals are generated as pressure waves and measurements are often pressure waves as well. In such applications, this formulation gives us a direct handle to integrating those physically relevant quantities into our computational model. This also makes it easier to link our model to the widely used (even though over-simplifying) acoustic approximation, which still represents one of the most popular computational models for seismic FWI. A direct link to those standard acoustic models can be established by letting the shear modulus in our model approach zero, \( \mu \to 0 \), which yields in the limit \( \epsilon_{xx} = \epsilon_{yy} = \epsilon_{xy} = 0 \). This provides us with a representation of the pressure-velocity model for the acoustic wave equation [35]. Many seismic wave simulators still rely on numerical implementations of the corresponding second-order scalar acoustic wave equation. Therefore, our approach can be considered a direct extension of this more classical simplified formulation, since it explicitly contains all acoustic quantities in the formulation. With this property, it also seems easier in future research to compare or even combine both approaches in an efficient hybrid inversion scheme.

3. The inverse problem

3.1. Formulation of the inverse problem

Let us now analyse in some more details the general structure of the system (17)-(22), and its impact on solution strategies for the inverse problem. In order to obtain a more compact notation, we define the tensors

$$
\Gamma(m) = \begin{bmatrix}
\rho & \rho & 0 \\
1/2\mu & 1/2\mu & 0 \\
0 & 1/\mu & 1/\lambda
\end{bmatrix} = \text{diag}(\rho, \rho, 1/2\mu, 1/2\mu, 1/\mu, 1/\lambda),
$$

$$
D_x = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\quad \text{and} \quad
D_y = \begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}.
$$

Here we use the standard notation \( \text{diag}(\ldots) \) denoting a diagonal tensor with the indicated elements \( \ldots \) on the main diagonal. Now the above system can be written in form of a symmetric hyperbolic system [35]

$$
\Gamma(m)w_t + D_x w_x + D_y w_y = q
$$
where \( q = (F_x, F_y, 0, 0, 0)^T \) is the source term; \( m = (\lambda, \mu, \rho) \) contains the material parameters governing the elastic wave propagation; \( w = (v_x, v_y, \epsilon_{xx}, \epsilon_{yy}, \epsilon_{xy}, p)^T \) contains the dynamic variables. By defining furthermore the elastic wave operator

\[
\mathcal{L}(m) = \Gamma(m)\partial_t + D_x\partial_x + D_y\partial_y,
\]

and adding appropriate initial and boundary conditions, we obtain the formulation of the forward problem

\[
\begin{align*}
\mathcal{L}(m)w &= q \quad \text{(27)} \\
w(x, 0) &= 0 \quad \text{(28)} \\
w(\partial\Omega, t) &= 0 \quad \text{(29)}
\end{align*}
\]

for \( x \in \Omega \) and \( t \in [0, T] \). Here, \( \Omega \) represents the survey domain and \([0, T]\) the observation time of the physical experiment. Notice that the operator \( \mathcal{L} \) is skew-self-adjoint for appropriately chosen function spaces in the sense that \( \mathcal{L}^* = -\mathcal{L} \), where \( \mathcal{L}^* \) is the formal adjoint (or dual) to \( \mathcal{L} \). As we will see further below, \( \mathcal{L}^* \) is associated with the solution of the following adjoint (or dual) symmetric hyperbolic system

\[
\begin{align*}
\mathcal{L}^*(m)z &= q_a \quad \text{(30)} \\
z(x, T) &= 0 \quad \text{(31)} \\
z(\partial\Omega, t) &= 0 \quad \text{(32)}
\end{align*}
\]

where the right-hand side \( q_a \) will conveniently be chosen later in \( \Theta \). The initial condition \( \text{(31)} \) and boundary condition \( \text{(32)} \) follow from applying integration by parts in the derivation of the formal adjoint equation. This adjoint system will play a fundamental role when designing the reconstruction algorithms of this paper. The formal derivation of adjoint schemes for inverse wave problems is by now well-understood such that we refer for most of the further discussions to the literature, in particular \( \text{[32, 6, 35, 37, 2, 38, 11, 5]} \). Here we want to highlight that in our particular formulation \( \text{(26)} \), essentially a minus sign distinguishes the forward from the adjoint operator. This reflects the fact that the forward problem is physical and causal, but that the adjoint problem corresponds to non-physical ‘time-reversal’ or ‘back-propagation’: adjoint residual signals are propagating backwards in time and space through the elastic medium in order to calculate distributed ‘sensitivities’ with respect to mismatches in the data. Accordingly, in the formulation of the adjoint problem, final value conditions need to be specified in \( \text{(31)} \) instead of the initial value conditions of the forward problem. Moreover, the forward and the adjoint problem formulations both assume that in finite time (the time of the experiment \([0, T]\)) no wave energy reaches the boundary \( \partial\Omega \), which is reflected by the conditions \( \text{(29)} \) and \( \text{(32)} \). This can always be achieved by assuming that energy enters the domain only by the given localized sources (which in the adjoint formulation are positioned at the receiver locations, see further below) and then putting the (freely chosen) boundary \( \partial\Omega \) sufficiently far away from those sources and receivers.

Denote by \( S \) the index set of sources and by \( R \) the index set of receivers. Furthermore, denote by \( d^{\text{cal}}_{\text{obs}}(x, r, t) \) the data observed at receiver \( r \in R \) located at \( x \), due to source \( q^s \), \( s \in S \). Furthermore, denote by \( [d^{\text{cal}}_{\text{obs}}(m)](x, r, t) \) those data calculated (predicted) given a model \( m \). For simplicity of the derivation, we assume that all fields \( w^s \), \( s \in S \), are functions of a standard \( L^2 \)-Hilbert space \( U \) equipped with an appropriate inner product \( \langle ., . \rangle \). Similarly, we can formally denote a Hilbert space of sources \( q^s \) by \( Y \). Under this assumption, when plugging sources \( q^s \in Y \) into \( \text{(27)} - \text{(29)} \), the resulting fields \( w^s \) can be modelled mathematically as functions in the standard Hilbert space \( U = L^2(X)^6 \) with \( X = \Omega \times [0, T] \). We want to mention, however, that the formal assumption of standard \( L^2 \)-Hilbert spaces can be generalized in our derivation by using appropriate duality pairs of spaces and their corresponding dual spaces instead, see for example \( \text{[39, 38]} \) for details. For that reason, we will not specify explicitly in the following the particular spaces used when writing inner products \( \langle ., . \rangle \) (or corresponding duality mappings). It will always be clear from the context which spaces are used. In a similar way, we assume that measurements are modelled by suitable measurement operators \( P_R : U \to Z \), where \( Z \) represents the function space of time traces \( d^{\text{obs}}_{\text{obs}}(x, r, t) \) at the receiver positions \( x \). In particular, we assume that \( P_R(w^s(m)) = d^{\text{cal}}_{\text{obs}}(m) \). The specific nature of those measurements is not of importance for our general derivation of the inversion scheme, and the specific nature of \( Z \) depends on details of the measurement process, including the number of independent components gathered in a field experiment. They also might vary with the specific source index \( s \). For simplicity of the notation, and without any restriction of generality, we assume in the theoretical analysis of this chapter that \( P_R \) is linear and that \( Z \) is independent of \( s \). In our numerical experiments, the specific details of this operator will be adjusted as necessary, following simple rules of assigning adjoint (or dual)
maps $P_R^*$ to the chosen measurement maps $P_R$. We also mention that often it is convenient to combine all $w^s$, $s \in S$, into a vector $w$, all sources $q^s$ into a vector $q$, and all data into a vector $d$.

A standard approach is now to formulate the least-squares data misfit functional

$$ J(m) = \frac{1}{2} \sum_{s \in S} \|d^s_{\text{cal}}(m) - d^s_{\text{obs}}\|_2^2 $$

(33)

and formulate the inverse problem as a constrained optimization problem

$$ m^* = \arg\min J(m) \text{ subject to the constraints} $$

$$ P_R^* w^s = d^s_{\text{cal}} $$

(35)

$$ L(m) w^s = q^s $$

(36)

One possible way of addressing the inverse problem is to formally eliminate the constraints from the problem by writing in (33)

$$ d^s_{\text{cal}}(m) = P_R^* [L(m)]^{-1} q^s $$

(37)

which yields a corresponding unconstrained formulation of the problem

$$ m^* = \arg\min \frac{1}{2} \sum_{s \in S} \|P_R^* [L(m)]^{-1} q^s - d^s_{\text{obs}}\|_2^2 $$

(38)

However, we will formally follow here an alternative approach that addresses the constrained formulation directly. We describe it in the following subsection.

### 3.2. Calculation of descent directions

In order to calculate formal expressions for descent directions, we follow here the general Lagrangian approach for constrained optimization, as outlined for example in [38, 6]. We mention that a perturbation approach can be taken as well which yields equivalent expressions. For details on that alternative approach, see [38, 40, 41, 42, 43].

Let us introduce a Lagrangian dynamic state vector $z$, which has the same length as $w$, and the associated Lagrangian data vector $\zeta$ which has the same length as $d$. Then we can formulate the Lagrangian as

$$ L(w, m, d_{\text{cal}}, z, \zeta) = J(d_{\text{cal}}, m) - \sum_{s \in S} \langle z^s, L(m) w^s - q^s \rangle - \sum_{s \in S} \langle \zeta^s, P_R^* w^s - d^s_{\text{cal}} \rangle. $$

(39)

Formal first-order conditions for a solution of the constrained problem are (for all $s \in S$)

$$ \frac{\partial L}{\partial w^s} = 0 \text{ or } L^s(m) z^s = -P^s_R^* \zeta^s; $$

(40)

$$ \frac{\partial L}{\partial z^s} = 0 \text{ or } L(m) w^s = q^s; $$

(41)

$$ \frac{\partial L}{\partial \zeta^s} = 0 \text{ or } P_R^* w^s = d^s_{\text{cal}}; $$

(42)

$$ \frac{\partial L}{\partial d_{\text{cal}}} = 0 \text{ or } d^s_{\text{cal}} - d^s_{\text{obs}} = -\zeta^s; $$

(43)

$$ \frac{\partial L}{\partial m} = 0 \text{ or } \frac{\partial J}{\partial m} = \sum_{s \in S} \langle z^s, \frac{\partial L}{\partial m} w^s \rangle. $$

(44)

In the fourth condition (43) we have used that $\frac{\partial J}{\partial m_{\text{cal}}} = d^s_{\text{cal}} - d^s_{\text{obs}}$. The symmetric hyperbolic elastic scheme (26) implies that formally $\frac{\partial J}{\partial m_{\text{cal}}} = \frac{\partial J}{\partial d_{\text{cal}}}$, such that the fifth condition (44) gives us

$$ \frac{\partial J}{\partial m} = \sum_{s \in S} \langle z^s, \frac{\partial J}{\partial d_{\text{cal}}} \rangle. $$

(45)

Combination of those conditions (40)-(45) will provide us with the building blocks of a gradient-based optimization scheme using efficient forward and adjoint solutions of the given symmetric hyperbolic system. As already mentioned,
a very useful feature of this formulation turns out to be that \( L \) is skew-self-adjoint [36 32], such that the same numerical code can be used for the required forward simulations in (41) and adjoint simulations in (40) with initial and boundary conditions properly adjusted. Notice also that the right-hand side of the adjoint system in (40) is further specified in (45), which yields

\[
L'(m)z^s = P_R^s(d_{cal}^s - d_{obs}^s) = q_a^s. \tag{46}
\]

Its physical interpretation is as follows. In order to computationally obtain a descent direction for the constrained optimization problem (43)-(46), several steps are required. First, the non-zero residuals at the receivers (assuming that the algorithm has not yet converged) are applied as artificial adjoint sources \( q_a \) (via adjoint projection \( P_R^s \)) at those receiver positions. They are then back-propagated into the medium by solving the adjoint system (30)-(32) in a time-reversed fashion. The condition (44) (or (45)) will finally indicate where and how to extract gradient-based updates from the resulting forward (\( w^s \)) and adjoint (\( z^s \)) wave fields for the individual material parameters in order to reduce the mismatch in the data further.

The specific gradient directions for each of the medium parameters are easily obtained from (45) due to the diagonal structure of \( \Gamma \). We need to calculate the partial derivatives with respect to the three elastic parameters \( \lambda, \mu, \rho \) which appear in \( m \). We identify in (23) contributions for each of those parameters by explicitly stating the partial derivatives

\[
\frac{\partial \Gamma}{\partial \lambda} = \text{diag} \left( 0, 0, 0, 0, 0, -1/\lambda^2 \right), \tag{47}
\]

\[
\frac{\partial \Gamma}{\partial \mu} = \text{diag} \left( 0, 0, -1/2\mu^2, -1/2\mu^2, -1/2\mu^2, 0 \right), \tag{48}
\]

\[
\frac{\partial \Gamma}{\partial \rho} = \text{diag} \left( 1, 1, 0, 0, 0, 0 \right). \tag{49}
\]

Following as before the notation of state variables \( w = (v_x, v_y, \epsilon_{xx}, \epsilon_{yy}, \epsilon_{xy}, p)^T \), and denoting in a similar fashion adjoint state variables as \( z = (v_x^*, v_y^*, \epsilon_{xx}^*, \epsilon_{yy}^*, \epsilon_{xy}^*, p^*)^T \), we obtain then expressions for all three partial derivatives of the gradient from (45) as follows

\[
\frac{\partial \mathcal{J}}{\partial \lambda} = \sum_{s \in S} \langle z^*, \frac{\partial \Gamma}{\partial \lambda} \frac{\partial w^s}{\partial t} \rangle_U = -\frac{1}{\lambda^2} \sum_{s \in S} \int_0^T \left( \langle p^s \rangle_t \frac{\partial p^s}{\partial t} \right) \, dt, \tag{50}
\]

\[
\frac{\partial \mathcal{J}}{\partial \mu} = \sum_{s \in S} \langle z^*, \frac{\partial \Gamma}{\partial \mu} \frac{\partial w^s}{\partial t} \rangle_U = -\frac{1}{2\mu^2} \sum_{s \in S} \int_0^T \left( (\epsilon_{xx}^s)_{,t} + (\epsilon_{yy}^s)_{,t} + 2(\epsilon_{xy}^s)_{,t} \right) \, dt, \tag{51}
\]

\[
\frac{\partial \mathcal{J}}{\partial \rho} = \sum_{s \in S} \langle z^*, \frac{\partial \Gamma}{\partial \rho} \frac{\partial w^s}{\partial t} \rangle_U = \sum_{s \in S} \int_0^T \left( (v_x^s)_{,t} \frac{\partial v_x^s}{\partial t} + (v_y^s)_{,t} \frac{\partial v_y^s}{\partial t} \right) \, dt. \tag{52}
\]

In FWI applications (including our numerical experiments), often a different set of physical parameters is considered, which replaces the Lamé parameters \( \lambda \) and \( \mu \) by p-wave and s-wave velocities \( V_p \) and \( V_s \), respectively (keeping \( \rho \) in both formulations). The transformation and back-transformation are done according to

\[
\lambda = \rho (V_p^2 - 2V_s^2), \quad \mu = \rho V_p^2, \quad \rho = \rho, \quad V_p = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad V_s = \sqrt{\frac{\mu}{\rho}}. \tag{53}
\]

Gradient expressions of both formulations are directly related to each other by the chain rule. When using \( V_p \) and \( V_s \) as unknown parameters of the inverse problem, the corresponding least-squares cost will be denoted here by \( \mathcal{J} \), and the gradient expressions are obtained (using the chain rule) as

\[
\frac{\partial \mathcal{J}}{\partial V_p} = 2\rho V_p \frac{\partial \mathcal{J}}{\partial \lambda}, \tag{54}
\]

\[
\frac{\partial \mathcal{J}}{\partial V_s} = -4\rho V_s \frac{\partial \mathcal{J}}{\partial \lambda} + 2\rho V_p \frac{\partial \mathcal{J}}{\partial \mu}, \tag{55}
\]

\[
\frac{\partial \mathcal{J}}{\partial \rho} = (V_p^2 - 2V_s^2) \frac{\partial \mathcal{J}}{\partial \lambda} + V_s^2 \frac{\partial \mathcal{J}}{\partial \mu} + \frac{\partial \mathcal{J}}{\partial \rho}. \tag{56}
\]
In the remaining parts of this paper, we drop the tilde on top of the symbol $\mathcal{J}$ in our notation and always use the appropriate formulation, which will not cause any confusion.

4. A level set formulation for seismic FWI

4.1. Level set representation of geological regions

For the mathematical representation of bodies of salt embedded in the domain of interest, we formally assume that the computational domain $\Omega$ is composed of two disjoint regions of significantly different characteristics. One of those regions is called $D \subset \Omega$, which is the region filled with some material of special interest. For simplicity in notation we assume that $D$ is (topologically) closed with boundary $\Sigma = \partial D \subset D$. The surrounding area is defined as the region $D^c = \Omega \setminus D$ and is assumed to be filled with approximately known background material. The region $D$ can have a complicated topology, consisting for example of several disconnected subregions of quite arbitrary shapes; Numerically, we can represent such a scenario conveniently by defining a (sufficiently regular) real-valued level set function $\phi(x)$ with the property that

$$
\begin{align*}
\phi(x) &\geq 0 \quad \text{for} \quad x \in D, \\
\phi(x) &< 0 \quad \text{for} \quad x \in D^c.
\end{align*}
$$

See the right hand image of Figure 4 for an illustration of this implicit representation of shapes. For more details of the level set approach for representing shapes in a computational environment we refer to \cite{26, 27, 28, 41, 44, 45, 46, 47}. In our numerical experiments, $D$ will represent the unknown salt domes, and $D^c$ represents the background lithology into which the salt domes are embedded. We can therefore write

$$
\begin{align*}
m(x) &= m_{obj}(x) \quad \text{where} \quad \phi(x) \geq 0, \\
m(x) &= m_{back}(x) \quad \text{where} \quad \phi(x) < 0.
\end{align*}
$$

Formally, all three functions $m, m_{obj}$ and $m_{back}$ are defined on the entire domain $\Omega$. The sign of the level set function $\phi$ decides which profile to use at a given location. Using the notation of a characteristic function $\chi_D$, such that $\chi_D(x) = 1$ for $x \in D$, and $\chi_D(x) = 0$ for $x \in D^c$ the model can also be written in the form $m = m_{obj}(x)\chi_D(x) + m_{back}(x)(1 - \chi_D(x))$. Alternatively, when using instead an appropriately defined one-dimensional Heaviside function $H$, we can identify $H(\phi(x)) = \chi_D(x)$. With this notation, we have

$$
m = m_{obj}(x)H(\phi) + m_{back}(x)(1 - H(\phi)).
$$

For a sufficiently smooth level set function $\phi$ we can identify furthermore $\Sigma = \{ x : \phi(x) = 0 \}$. Let us assume for the moment that the profiles $m_{obj}(x)$ and $m_{back}(x)$ are provided to us (we will later relax this assumption), and that we only have to identify from the collected data the (possibly complicated) interface $\Sigma$ of the regions $D$ and $D^c$ separating the regions filled with the different profiles. In a shape-evolution-type optimization approach, we would aim to design an evolution of the representing level set function over an artificial pseudo-time $\tau \in [0, \Theta]$ with a (still to be chosen) stopping time $\Theta > 0$. This is done by writing $\phi = \phi(x, \tau)$ satisfying the corresponding Hamilton-Jacobi type evolution equation

$$
\begin{align*}
\frac{\partial \phi}{\partial \tau} + V(x, \tau) \cdot \nabla \phi &= 0, \\
\phi(0) &= \phi(0),
\end{align*}
$$

on $\Omega \times [0, \Theta]$. Here, $V$ is a suitably designed (vector) velocity field which ideally points into a descent direction to a given data misfit functional at a particular (pseudo-)time step $\tau$. For obtaining such an evolution, it is sufficient to consider the evolution scheme \cite{26, 27, 28}

$$
\begin{align*}
\frac{\partial \phi}{\partial \tau} + F(x, \tau)|\nabla \phi| &= 0, \\
\phi(0) &= \phi(0),
\end{align*}
$$

on $\Omega \times [0, \Theta]$ where $F = V \cdot n$ and $n = \frac{\nabla \phi}{|\nabla \phi|}$ is the corresponding outward normal direction. After time-discretization (e.g. using a forward-Euler scheme) $\tau_{n+1} = \tau_n + (\Delta \tau)n$, we arrive at an optimization approach of the form

$$
\phi_{n+1} = \phi_n - \alpha_n F_n|\nabla \phi_n|.
$$
with \( \alpha_n = (\Delta \tau)_n \) being an appropriately chosen step length which usually depends on the iteration index \( n \), and \( \phi_n = \phi(x, \tau_n) \), \( F_n = F(x, \tau_n) \). Our goal is therefore to derive efficient choices for \( \alpha_n \) and \( F_n \) such that the above scheme converges to a solution of our constrained optimization problem. The selection of \( F_n \) requires us to derive appropriate descent directions for a given cost functional, whereas the selection of \( \alpha_n \) amounts to designing efficient line search criteria.

Formally, descent directions can be obtained from our previous discussion by applying the chain rule to the modified cost functional (now depending on the level set function \( \phi \) by function composition)

\[
\phi^* = \arg\min \mathcal{J}_{ls}(\phi) \quad \text{with} \quad \mathcal{J}_{ls}(\phi) = \frac{1}{2} \sum_{s \in S} \|d_{cal}^s(m(\phi)) - d_{obs}^s\|^2,
\]

under the constraints provided by \( (35), (36) \). Taking into account also the pseudo-time dependence of \( \phi \), we arrive by the chain rule at

\[
\frac{d\mathcal{J}_{ls}}{d\tau} = \frac{df}{dm} \frac{dm}{d\phi} \frac{d\phi}{d\tau} = \left( (m_{obj} - m_{back}) \delta(\phi) \right) \frac{d\phi}{d\tau} \cdot g_m
\]

where \( g_m \) denotes the gradient of \( \mathcal{J} \) with respect to the medium parameter \( m \) which is given by \( (59) - (62) \) in the \( (\lambda, \mu, \rho) \) formulation or \( (54) - (56) \) in the \( (V_n, V_s, \rho) \) formulation. The Dirac delta \( \delta(\phi) \) enters here as the formal derivative of the Heaviside function \( H(\phi) \) in \( (61) \) and provides the formal link between our optimization approach and the concept of a shape evolution: Only field values at the current interface contribute to the calculation of the descent direction, since only points close to this interface are affected by a formal shape evolution at a fixed pseudo-time step. Since we know already how to calculate \( g_m \), we can use \( (66) \) for determining descent directions of \( \mathcal{J}_{ls} \) with respect to the level set evolution \( \phi(\tau) \) at each pseudo-time step \( \tau_n \). For more details on such technical aspects of applying level set approaches to solving inverse problems, we refer here to the recent survey papers \[44, 45, 46, 47\]. In the following, we will mention two convenient aspects which help to stabilize our inversion scheme for the application of seismic FWI, namely the use of a narrow-band for gradient localization and the use of Sobolev spaces for regularization.

### 4.2. Gradient localization by a narrowband approach

The evolution law \( (66) \) indicates that descent directions are determined by evaluations of field values (via \( g_m \)) located at the points \( x \) along the current interface between \( D(\tau) \) and \( D'(\tau) \) satisfying \( \phi(x, \tau) = 0 \). This has the effect that information of the (forward and adjoint) fields is used only along the current interface for finding descent directions for \( \mathcal{J}_{ls} \). This has interesting consequences on the overall optimization procedure, since the search for the minimum is restricted to follow a constrained search path. This can have advantages as well as disadvantages in the application of seismic FWI. As a potential drawback, restricting the parameter evolution by additional constraints has the potential to introduce specific local minima. However, this can be alleviated by choosing a convenient initial guess for the shape evolution. On the other hand, localizing the evolution to subregions of the domain can have beneficial effects when working in so-called low-sensitivity regions. These are regions where perturbations of medium parameters only have a very small impact on the data for the given source-receiver arrangements. In seismic FWI, these are mainly regions located deeper inside the Earth. Localizing gradient evaluation to those regions helps identifying those objects which are located in deeper low-sensitivity zones.

Computationally, it is difficult to implement an exact Dirac delta function (or distribution) in particular when using a finite size computational grid. Moreover, the level set evolution law \( (63) \) requires us to provide a velocity function \( F(x, \tau) \) not only along the current interface \( \Sigma \) but everywhere in the domain. Practically, both issues are addressed by defining a suitable ‘extension velocity’ which can be designed according to the needs of the application. In particular, this gives us some choice for designing the shape evolution some distance away from the current interface. Different strategies have been proposed in the literature for a variety of different situations, for specifying such extension velocities. As part of an optimization scheme as ours, a balance needs to be found between the stability of the evolution itself and the overall goal of arriving in an efficient way to a (local or global) minimum of the underlying optimization approach. We have made good experience with the employment of a so-called narrow-band function \( \chi_{nbd(\phi, \delta)}(x) \) which defines a descent direction by setting

\[
\frac{d\phi}{d\tau} = -(m_{obj} - m_{back}) g_m \chi_{nbd(\phi, \delta)}.
\]

Here, \( \chi_{nbd} \) indicates the characteristic function concentrated on a narrow neighbourhood-band around the interface \( \Sigma \) of a prescribed width, the so-called narrow-band, which is formally approximating the Dirac delta in \( (66) \). See
the right image of Figure 4 for an illustration of this situation. In fact, by applying a Sobolev gradient scheme as explained further below, this narrowband is furthermore slightly smoothed out towards its neighbourhood in order to provide us with a sufficiently smooth approximation of the Dirac delta and therefore a smooth level set evolution overall. Notice that the descent property of (66) remains (by all practical means) conserved by these modifications to arrive at an extension velocity. The narrowband approach makes sure that updates remain sufficiently local in a small neighbourhood of the current interface (thereby remaining close to the concept of a ‘shape evolution’ and helping to explore low-sensitivity regions) but makes sure that the evolution is well-defined and meaningful in the entire domain Ω. After time-discretization we arrive at the iteration rule

\[ \phi_{n+1} = \phi_n - \alpha_n g_\phi(\phi_n) \]  

(68)

with

\[ g_\phi(\phi_n) = -(m_{\text{obj}} - m_{\text{back}})g_m\chi_{\text{nbd}}(\phi_n, d). \]  

(69)

We mention that the step length \( \alpha_n \) (also in the following) will be chosen in order to fix the number of pixels (let us call this quantity \( \nu \)) that change value between \( m_{\text{obj}} \) and \( m_{\text{back}} \) (or vice versa) in a given update to stay between a lower bound \( \nu \) and an upper bound \( \bar{\nu} \) such that \( \nu > \nu > \bar{\nu} > 0 \). Practically, this is done by a straightforward backtracking line search on \( \alpha_n \) which does not require us to run any additional forward or adjoint simulations and therefore comes without any significant computational expense. For more details, see [48] [49] [50] [51].

4.3. Sobolev spaces for regularization

The underlying inverse problem is highly ill-posed for various reasons. In addition to the high non-linearity of \( \mathcal{F} \) with respect to the unknown medium parameters, which potentially gives rise to a significant number of local minima (‘pseudo-solutions’), also the limited aspect ratio of sources and receivers and the limited frequency content of the probing fields can produce instabilities during the inversion. Therefore, regularization is a fundamental aspect of any inversion approach for seismic full waveform data. The use of a shape-based model can be considered a line search on \( \alpha_n \), which does not require us to run any additional forward or adjoint simulations and therefore comes without any significant computational expense. For more details, see [48] [49] [50] [51].

We will investigate the use of a Sobolev metric or norm for this purpose [52] [39], in two different ways. Let us consider (sufficiently smooth) level set functions \( \phi \in L^2(\Omega) \). For any two of such functions \( \phi_1 \) and \( \phi_2 \), we denote the inner product

\[ \langle \phi_1, \phi_2 \rangle_{W_\gamma(\Omega)} = \langle \phi_1, \phi_2 \rangle_{L^2(\Omega)} + \gamma \langle \nabla \phi_1, \nabla \phi_2 \rangle_{L^2(\Omega)} \]  

(70)

with appropriately chosen parameter \( \gamma > 0 \) and the associated norm \( ||w||_{W_\gamma(\Omega)}^2 = \langle \phi, \phi \rangle_{W_\gamma(\Omega)} \). Here, \( \langle \ldots \rangle_{L^2(\Omega)} \) refers, as before, to the standard \( L^2 \)-inner product for the space of wavefield components [39]. The space of functions \( \phi \) with finite \( ||\phi||_{L^2(\Omega)} \) (and appropriate behaviour towards the boundary of \( \Omega \)) is called here \( W_\gamma(\Omega) \). We will test the use of this space in our numerical experiments in two different ways.

First, notice that the gradient direction of \( \mathcal{F} \) with respect to the level set function \( \phi \) actually depends on the space where \( \phi \) is supposed to live in, which is reflected by the inner product used for calculating adjoint operators. The so-called Sobolev gradient direction assumes that \( \phi \in W_\gamma(\Omega) \) and thereby all available update directions need to be as well in this space. On the other hand, the gradient direction \( g_\phi \) derived in [60] assumes \( L^2 \)-spaces for the underlying functions and inner products, which allows for functions and updates with less smoothness properties. It is shown in [53] [48] [54] that the corresponding Sobolev gradient \( g_{\phi, \gamma} \) can be obtained from the previously determined \( L^2 \) gradient \( g_\phi \) by a simple post-processing scheme

\[ g_{\phi, \gamma} = (I - \gamma \Delta)^{-1} g_\phi \]  

(71)

where the filtering operator \( (I - \gamma \Delta)^{-1} \) is supplemented with appropriate boundary conditions at \( \partial \Omega \). \( I \) denotes the identity operator and \( \Delta \) is the Laplace differential operator. Different choices for \( \gamma \) provide gradients (and thereby
updates in each iteration) with varying degree of smoothness. We will employ Sobolev gradients in all numerical experiments shown further below. In particular, we will perform the modified updates

\[ \phi_{n+1} = \phi_n - \alpha_n \nabla J(\phi_n) \] (72)

instead of (68).

A second possible use of Sobolev norms is provided by adopting the more classical approach for regularization based on adding a Tikhonov term to the least-squares data misfit. By doing so we obtain the regularized cost functional

\[ J_{\theta,\gamma}(\phi) = J(\phi) + \theta \| \phi \|^2 \] (73)

with \( \theta \geq 0 \). When choosing \( \theta = 0 \) we formally have the same situation as before. However, when choosing \( \theta > 0 \), an additional term is present in the cost functional which changes the overall behaviour of the algorithm. More specifically, the attempt of minimizing \( J_{\theta,\gamma}(\phi) \) instead of \( J(\phi) \) gives rise to additional contributions to the updates which are analogous to those derived above, but are applied directly to the level set function as opposed to the updates \( \phi_n \). The resulting update rule is called

\[ \phi_{n+1} = \phi_n - \alpha_n g_{\theta,\gamma}(\phi_n) \] (74)

which is identical to (72) if \( \theta = 0 \) but performs additional smoothing if \( \theta > 0 \). This additional smoothing can have beneficial but also unwanted effects on the reconstructions. On the one hand, the added term provides an additional contribution to smoothing out the reconstructions, which overall helps removing small unwanted artefacts. On the other hand, we recall that in FWI a given source-receiver setup will have sensitivities towards parameter perturbations inside the earth which strongly depend on the location of a considered point and usually decrease with depth. This means in particular that inaccuracies in the model at deeper layers of the earth will only have a very small impact on the data. The additional term in (73) could easily dominate those contributions such that structures located deeper in the earth might be removed or not reconstructed at all by minimizing (73), whereas they might still be recovered when considering (55) instead. We will come back to this observation when discussing our numerical experiments in more details.

4.4. A nonlinear Kaczmarz scheme for data processing

We have above derived expressions for calculating descent directions for the least-squares functional \( J \) efficiently using an adjoint scheme. The most time-consuming part of this is the simulation of one forward and one adjoint seismic wave propagation problem for each seismic source using the elastic scheme. Therefore, with this setup, as many forward and adjoint simulations are required for the calculation of the full (vector-) gradient of the problem as there are seismic sources applied. This number can easily amount to hundreds or even thousands. Even though the calculation of the full gradient might enable us to employ many of the more traditional gradient-based optimization approaches (such as for example quasi-Newton or nonlinear conjugate gradient schemes [56]), each individual step of such iterations would still be extremely time and memory consuming. In particular, at an early stage of the shape evolution, where we are still far away from the actual shape, this would be highly inefficient. Therefore, we employ an alternative nonlinear Kaczmarz type approach, where descent information for each individual data set is used as soon as it has been calculated.

To be more precise, let us assume here that the entire set of sources \( S \) consists of \( K \) different source positions, which we denote \( s_k, k = 1, \ldots, K \). In the nonlinear Kaczmarz type approach, we divide \( S \) into subsets which are dealt with individually, one after the other. In the sequential approach employed in our numerical experiments, for example, we will simply define \( S_k = \{ s_k \}, k = 1, \ldots, K \), each containing just the source position with index \( k \). Then we apply the above analysis regarding descent directions to \( S_k \) instead of \( S \), providing us with descent directions in each step \( n \) of the algorithm which only are designed to reduce the data misfit with respect to the chosen source position. A ‘sweep’ of our algorithm is completed after all source positions are considered exactly once in this sequential fashion. We will perform usually many such sweeps in order to obtain improved accuracy. Let us use the index \( l = 1, \ldots, L \) for numbering the sweeps. Then the overall iteration index \( n \) in sweep \( l \) dealing with source \( s_k \) can be written as \( n = (l - 1)K + k \).

This Kaczmarz type approach is very efficient since it only requires us to look at one source at a time. However, we need to design new strategies to determine optimal step sizes in each iteration, and we need to agree on a quality measure for monitoring progress and (optionally) deciding when to stop the iterations. Directly monitoring the values
of expression in (65) at each sweep is not very efficient since it would require us to calculate $K$ additional forward problems for each sweep number $l$. Instead, we decide to use only the information which automatically is obtained during our iterations by defining (and plotting) for each sweep number $l$ the scalar quantity

$$J_{ls}^{(l)} = \frac{1}{2} \sum_{k=1}^{K} ||d_{cal}^{s_k}(m(\phi_{(l-1)K+k})) - d_{obs}^{s_k}||_2^2.$$  

(75)

All expressions on the right-hand side are calculated before doing the update in step $n = n(k, l) = (l-1)K + k$, and are obtained when calculating the corresponding descent directions. Therefore, monitoring the quality measure (75) can be done without any significant computational cost. Comparisons of (75) with the expression in (65) (not displayed here) have shown that both measures agree very well with each other and that monitoring (75) instead of (65) is a very good and reliable quality measure during the reconstruction process, in particular at later iterations where the material parameters only slightly change in each individual update.

On multi-processor architectures, also block-Kaczmarz or other more sophisticated single-step approaches can be employed as well where the sets $S_k$ are chosen differently. The idea of those is to use partial data in each iteration of the optimization loop in the most efficient way possible. We mention that related techniques have become popular in many other applications, including the stochastic gradient descent (SGD) method in large scale machine learning applications [57, 58, 59, 60]. In some of those works it is reported that the Kaczmarz approach might also be efficient in avoiding to being trapped in some local minima. See also [39, 41, 42, 43, 51, 48, 54, 61, 53] for similar approaches.

4.5. Reconstruction of internal parameter values

So far we have assumed that the values $\lambda_{obj}$, $\mu_{obj}$, $\rho_{obj}$ are a-priori known and constant. Often, in the mapping of salt domes, good estimates for those values are available such that this assumption seems reasonable. However, in cases where those values are highly uncertain, it might be advantageous to estimate them jointly with the interfaces of the regions in the same optimization loop. This task will be considered in this section. We will see in our numerical experiments that there actually might be additional good reasons for including the search for those parameters into the optimization scheme. For example, modifying those internal parameter values throughout the evolution can help avoiding certain local minima which otherwise would arise in purely shape-driven optimization approaches.

Practically, it is not difficult to incorporate this additional feature into our shape evolution scheme. All what is needed is to derive, in addition to shape gradients, also gradients with respect to these additional terms and then integrating the corresponding updates in the optimization loop. For simplicity, we will assume here that the internal values are constant but unknown. The case of smoothly varying internal profiles can be dealt with in a completely analogous way, but does not seem to provide any advantages in the applications considered here. Gradient directions for those internal profiles can be obtained simultaneously with the shape gradients for shape evolution by employing the chain rule as outlined in [46, 47], without significant additional cost. In our particular setup, where those parameters are assumed to be real-valued constants, also a direct finite-difference evaluation remains a feasible choice for determining these descent directions [46].

Regardless of which method is chosen, let us denote the gradient (or search) directions for the three scalar variables $\lambda_{obj}$, $\mu_{obj}$, $\rho_{obj}$ by $g_{\lambda}$, $g_{\mu}$ and $g_{\rho}$, respectively. Then, as part of an optimization loop, we define $\beta_n$ to be a step length at iteration step $n$, start with the initial values $\lambda_{obj}^0$, $\mu_{obj}^0$, $\rho_{obj}^0$, and define the update rule

$$\lambda_{obj}^{n+1} = \lambda_{obj}^n - \beta_n g_{\lambda}(\phi_n)$$

(76)

$$\mu_{obj}^{n+1} = \mu_{obj}^n - \beta_n g_{\mu}(\phi_n)$$

(77)

$$\rho_{obj}^{n+1} = \rho_{obj}^n - \beta_n g_{\rho}(\phi_n)$$

(78)

Those updates can be done simultaneously or in a specifically designed order. In our numerical experiments, we have chosen to only update one of the three parameters in each shape evolution step, and cycling this way over the three different variables. Moreover, we start this additional internal optimization loop after an initial fixed number of shape evolution sweeps $L_1$ with $1 \leq L_1 \leq L$, since updates for the internal parameter values are only meaningful when the shapes have already converged up to a certain degree. In our numerical experiments, we choose $L_1 = L/2$ where the total number of sweeps $L$ is assumed to be an even number.

Notice that, due to the cycling nature of the updates, we need to keep the step length $\beta_n$ small enough in order to acknowledge that each of the three parameters can have contributions to the least-squares data misfit. We use an empirical rule for fixing this step length individually for each parameter.
4.6. The algorithm

A short summarizing overview of the resulting algorithm is provided next.

**Algorithm 1**: Level set inversion for seismic FWI

| Choose initial model $m^{(0)}$ and corresponding $\phi_0 = \phi^{(0)}$; |
| Choose initial contrast values $\lambda^0_{obj}, \mu^0_{obj}, \rho^0_{obj}$; |
| Select sets $S_k = \{s_k\}$, $k = 1, \ldots, K$; |
| Choose steering parameters of the algorithm (e.g. $\gamma, \alpha_n, \beta_n, \ldots$); |
| Choose total number of sweeps $L$ to be performed and $L_1$ to start search for internal values; |
| Set $N = L : K$; |
| for $l = 1 : L$ (iteration over sweeps) do |
| for $k = 1 : K$ (iteration over subsets $S_k$) do |
| Select sequential Kaczmarz iteration order $n = n(k, l) = (l-1)K + k$; |
| Solve forward problem (27)-(29) and adjoint problem (30)-(32) for $S_k$; |
| Calculate descent directions $g_\phi(\phi_n)$ or $g_{\phi, \gamma, \theta}(\phi_n)$ from (69)-(71); |
| Determine step size $\alpha_n$ and perform update (72) or (74), respectively; |
| if $l \geq L_1$ (criterion for update of internal values) then |
| Do update for one of $\lambda_{obj}, \mu_{obj}, \rho_{obj}$ per $S_k$ in sequential order, using (76)-(78); |
| end |
| end |
| end |
| Final reconstruction is $m(\phi_N)$ using $\lambda^N_{obj}, \mu^N_{obj}, \rho^N_{obj}$. |

5. Numerical experiments

5.1. Numerical setup of the forward and adjoint problem

The computational modelling of elastic wave propagation has a long history in geophysical applications. We refer to [34] and references therein where a variety of different approaches is discussed in details, together with their specific advantages and disadvantages in particular situations. In our numerical experiments shown below, the system (17)-(22) will be discretized using a pseudo-spectral method following the general guidelines presented in [34]. Notice that [34] implements a slightly different system similar to (12)-(16), such that specific modifications of the computational scheme are necessary, which are described in more details in [49]. As general features, a staggered grid is used in the spatial variables and a forward leap-frog finite-differences scheme in the time variable. A suitable Courant-Friedrichs-Lewy (CFL) condition is enforced in order to guarantee stability, and multi-axial perfectly matched layers are added to the computational boundaries in order to avoid artificial reflections. The numerical scheme has been validated against results obtained by the open-source k-wave code [34]. We emphasize again that, due to the mentioned simple relationship between the forward and the adjoint model in our formulation, essentially the same numerical implementation is used in order to simulate back-propagation by the time-reversal adjoint system (30).

In order to test the proposed inversion algorithm, we consider synthetic 2D earth models as shown in Figures 2 and 13. The figures show the distribution of the three material parameters $V_p$, $V_s$ and $\rho$. The situation displayed in Figure 2 and used in Experiments 1-5 consists of a linearly varying background and some embedded objects representing for example salt domes. The objects share a common interface $\Sigma$ against the background in all three material parameters. The values chosen are listed in Table 1. The notation $[a, b]$ in the first row of the table indicates that the parameter changes linearly between the lower value $a$ and the upper value $b$. The second row lists the true internal parameter values inside the embedded objects, which are assumed to be constant but in general unknown. Instead, in the reconstruction algorithm the values listed in the third row of the table are applied at the initial guess inside the evolving shapes until they are updated as well according to calculated descent directions. The situation in Experiment 6 is shown in Figure 13 and is slightly more complex. It will be discussed in more details further below.

The starting guess for the shape evolution can in principle be arbitrary, even though certain configurations seem to work better in terms of avoiding local minima or speeding up overall convergence. Notice that the level set evolution as proposed here is a shape evolution which restricts updates to the boundaries of the current shapes, due to the
narrowband approach. See again the right image of Figure 4 for an illustration of this technique. Fast convergence might therefore more likely if initial shapes are either already reasonably close to the expected final shapes (as for example in Figures 5-12), or are evenly distributed (in a convenient pattern) over the entire domain, as in Figures 13, 15. We also provide an example in Figures 3, 4 where the initial guess is very small and localized. The reconstruction and shape evolution from such a small and isolated initial guess works well in those numerical experiments, but in general might be suboptimal due to expected long evolution time until convergence.

Table 1: parameter values of the test model in Experiments 1-5

<table>
<thead>
<tr>
<th></th>
<th>( V_p ) [m/s]</th>
<th>( V_s ) [m/s]</th>
<th>( \rho ) [kg/m³]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>[3000,4000]</td>
<td>[1200,2000]</td>
<td>[2100,2500]</td>
</tr>
<tr>
<td>True internal</td>
<td>5000</td>
<td>2700</td>
<td>3000</td>
</tr>
<tr>
<td>Initial internal</td>
<td>5500</td>
<td>3000</td>
<td>2700</td>
</tr>
</tbody>
</table>

The 3km \times 12km computational domain in 2D is discretized by a finite grid of size 30 \times 120 where each grid cell has a physical dimension of 100m \times 100m. 100 sources and 80 receivers are arranged along a straight line as shown in Figures 2 and 13. The red diamonds in the figures indicate source positions and the blue stars (inside the red diamonds) indicate receiver positions. In particular, the 80 source locations coincide with 80 of the 100 receiver positions, such that, at each side of the line, 10 receiver positions are located which are not source positions. The time-behaviour of the sources consists of Ricker wavelets of the form

\[
R(\tau) = (1 - 2\tau^2) \exp(-\tau^2), \quad \text{with} \quad \tau = \pi f(t - t_d) - k\pi.
\]

We apply the Ricker wave formulation in equation (79), and define as central frequency \( f = 1.5 \text{ Hz} \). Furthermore, \( t_d = 0 \text{s} \), \( k = 2 \). This Ricker wave satisfies a prescribed numerical dispersion condition.

The data are simulated with the above described forward modelling code. Then, 5% Gaussian noise is added to each data value individually. This is sufficient for our proof-of-concept purpose. For the future, we plan to generate data on a different grid size, or even use a different discretisation scheme for data generation and reconstruction. Gradient calculations are performed by using the same code as for the forward model, but invoking the time-reversal property for backward propagation of adjoint fields.

In our numerical experiments, we fix the total number of sweeps to 50. This fixed number of iterations is useful for comparing different reconstructions with each other. In experiments 2-6 we will reconstruct both, the shapes and the internal parameter profiles. As mentioned before, we will consider during the first 25 sweeps of the level set
evolution only the shape evolution without updating the internal parameters. Here we use the initial values provided in Table 1. At sweep 25 we then start the joint reconstruction of shapes and internal values. In those sweeps, we apply simultaneous updates for the shape and, alternating in a cyclic fashion, one of the three internal parameter value at the same time.

In the following subsections, we will present and discuss the results of our numerical experiments in order to demonstrate the performance of our algorithm in different situations. In some of the numerical experiments we will assume that the data consist of the full seismic record at the receiver locations. In other words, the measurement operator $P_R$ in (35) extracts all six components of the seismic wavefield. This is an idealized situation which often in practice is not achievable. Therefore, in another set of numerical experiments we restrict our measurements to only one component of the seismic wave field, namely pressure $p$ at the receiver locations. This is a more realistic situation in off-shore seismic situations where sources and receivers might be immersed in water where shear waves are not accessible.

We will consider in total six numerical experiments. They consider two different physical setups. The first setup is used in numerical experiments 1-5 and results are shown in Figures 2 to 12. Those are chosen in order to demonstrate some fundamental features of our algorithm in a well controlled situation. The second physical setup is used in numerical experiment 6 and results are shown in Figures 13 to 15 considering a slightly more complex situation. This particular experiment is chosen in order to test the sensitivity of our algorithm with respect to insufficient knowledge of the physical environment and its numerical stability in such cases.

Figure 2 shows the general setup of our first physical setup. In numerical experiments 1-5, we will show and com-
pare the performance of our algorithm when using different regularization schemes. In some of those experiments we will use Sobolev regularization without adding any extra term to the least-squares data misfit functional, as formulated in (33). In other experiments, an additional Tikhonov-Philips term is added to the least-squares data misfit functional as formulated in (73). We will compare the results of these different approaches with each other for some chosen regularization parameters.

5.2. Experiment 1: Some basic differences between shape-based and pixel-based schemes

To start with, in order to obtain a general understanding of the fundamental differences between shape-based evolution and pixel-based evolution, we first compare in Figures 3 and 4 the performance of a basic level set reconstruction with the performance of a standard Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS, see [56] for more details) quasi-Newton reconstruction scheme for the physical setup as shown in Figure 2. None of the two reconstruction schemes used in this initial comparison is designed to be highly sophisticated. They still provide an interesting comparison of the functioning of both fundamental reconstruction approaches: pixel-based on one hand, and shape-based on the other hand. In this particular setup, for comparison purposes, the level set reconstruction uses a Kaczmarz approach with correctly known internal parameter values of the shapes to be reconstructed. The starting guess consists of two very small objects positioned inside the domain. Since we aim at comparing both algorithms under similar conditions, the pixel-based scheme starts with the same initial guess as the level set reconstruction. This choice guarantees that the least-squares data misfit evolution starts out from the same numerical value in both schemes. It also highlights that the different schemes converge to different solutions, even though they start from the same initial guess.

Figure 3 shows on rows 1-3 the reconstruction of $V_p, V_s$ and $\rho$ for the pixel-based scheme, respectively. On rows 4-6 the corresponding reconstructions for the level set scheme are shown. In all rows, the columns one to five correspond to iteration numbers 1, 5, 10, 20 and 50, respectively. We mention that the aspect ratios of the displayed images in this figure is different from the one used in Figure 2 in order to fit all iterations on one row. The physical dimensions are still the same as in the original profile shown in Figure 2.

Figure 4 shows on the left the evolution of the least-squares data misfit for both schemes against iteration number. On the right the figure illustrates the concept of the narrowband in the level set approach, which localizes updates of the shape evolution around the boundaries of the current shapes. Notice that, as a consequence, the pixel-based scheme tends to provide volumetric updates of small contrast in each iteration, whereas the level set based scheme tends to provide updates of high contrast (given the internal parameter values) but spatially restricted to a small vicinity of the current shape boundary. Therefore, the small high-contrast initial shape objects will not be modified significantly by the pixel-based scheme and can still be seen in all iterations as small highlighted dots in the reconstructed profiles. The level set approach instead modifies those initial shapes locally, maintaining the prescribed high contrast, until the cost functional assumes a minimal value. It can be observed in particular that some (wave-like) oscillatory features are visible in the reconstructions of the pixel-based scheme at all iteration counts, which are not visible in the level set approach. Also, the least-squares data misfit value after 50 iterations is lower in the shape-based approach than in the pixel-based approach. This is understandable since we were using additional correct prior information in the shape-based approach.

5.3. Experiments 2-5: Further analysis of the level set scheme

Now we turn our attention to more detailed aspects of the level set reconstruction technique. The numerical results discussed next are shown in Figures 5-12. The initial guess for the level set function is chosen to be a signed distance function representing the two discs shown on the right-hand image of the top rows of Figures 5, 7, 9 and 11. All level set functions during the evolution, including this initial guess, are rescaled by multiplication with a scalar such that its maximum inside the estimated salt bodies takes the value 10. This is done in order to avoid very flat level set functions. The level set approach instead modifies those initial shapes locally, maintaining the prescribed high contrast, until the cost functional assumes a minimal value.

The remaining images in these figures show the reference shapes and snap shots at iteration numbers 10, 20 and 50 of of the reconstructions of $V_p, V_s$ and $\rho$. Figures 6, 8, 10 and 12, on the other hand, show the evolution of the internal parameter values as well as the least-squares data misfit functional during the 50 iterations. We mention that in the experiments 3 and 5, where actually the functional (73) is minimized, also the value for $J_l^{(i)}$ is displayed in the figures in order to facilitate the direct comparison of all obtained results.
Fig. 3: Numerical Experiment 1: The figure shows on rows 1-3, from left to right, the reconstructed pixel-based profiles of the L-BFGS method after 1, 5, 10, 20 and 50 iterations, respectively. First row: $V_p$, second row: $V_s$, third row: $\rho$. On rows 4-6 the figure shows, from left to right, the reconstructed profiles of a basic level set method after 1, 5, 10, 20 and 50 iterations, respectively. Row 4: $V_p$, row 5: $V_s$, row 6: $\rho$. 
It is clearly visible in the corresponding graphs displayed in the figures that the reconstruction of the internal parameter values is not started before sweep number 25. Until that sweep, only the shapes evolve according to the calculated descent directions with respect to the initially chosen internal parameter values. After sweep 25, internal parameter values are updated as well and descent directions are calculated for the shapes taking into account the latest best guess to the internal parameter values inside the salt bodies. We will see that this small modification at later sweeps can make an important difference for the convergence of our shape-based reconstruction scheme.

Figures 5 and 6 show the results for Numerical Experiment 2. It can be seen that the reconstruction after sweep 10 looks almost identical to the reconstruction after sweep 20. The data misfit is reduced mainly during the first three sweeps. Afterwards the evolution remains almost stationary. The search for the internal parameter values starts in sweep 25. Thereafter, the shape remains almost the same and only small changes occur in response to the changing internal parameter values. It can be seen that the three internal parameter values converge properly and the final values of all three elastic parameters are quite close to the correct reference values. We can observe also that the final shapes show several ‘small holes’ which build up during early iterations and do not disappear later on. They seem to represent some form of local minima, where these holes most likely are compensated by different features elsewhere in the reconstructed profiles.

Figures 7 and 8 show the results for numerical experiment 3. As mentioned, it differs from Experiment 2 by the added Tikhonov-Philips term in the data misfit functional which is used for the minimization algorithm. Again, the reconstruction after sweep 10 looks very similar to the one after sweep 20. However, at this stage the reconstruction shows significant missing parts compared to the correct reference model. Our interpretation of this is that the additional term in the cost functional limits the flexibility of the shape evolution in order to deform the shapes into the correct direction. In other words, the shape evolution arrived at a local minimum. However, once the reconstruction of the internal parameter values is started in sweep 25, also the shapes start evolving further accordingly. Our interpretation of this phenomenon is that by increasing the dimension of the parameter space, when adding the search for the internal parameters, the algorithm is able to escape the (shape-based) local minimum. The final shape after 50 sweeps is quite close to the correct reference shape. Notice also that in this reconstruction the final level set function is much smoother and more regular compared to the final level set function in experiment 1. This is due to the smoothing property of the added Tikhonov term, combined with the inherent Sobolev smoothing mechanism. The ‘small holes’, which were visible in the reconstruction of Experiment 2, practically have disappeared due to this additional smoothing throughout the shape evolution. Also the internal parameter values of the salt bodies are reconstructed quite well.

Based on these results so far, one wonders whether we should always add Tikhonov regularization to the shape reconstruction in order to arrive at smoother final shapes. We will argue in the next two experiments that adding such a term is not always a good idea.

Figures 9 and 10 show the same situation as Figures 5 and 6 using Sobolev smoothing but no additional Tikhonov term. However, now only one component of the elastic wavefield is used as data, namely the pressure. The wave
Fig. 5: Numerical Experiment 2: Only Sobolev regularization is applied and all six components of the elastic wave field are measured at receiver locations.
Fig. 6: Numerical Experiment 2: Only Sobolev regularization is applied and all six components of the elastic wave field are measured at receiver locations. The figure shows the evolution of the three internal parameter values as well as the least-squares cost functional.

Propagation itself still follows the complete elastic model including mode conversion and shear waves. The evolution overall follows the same pattern as in Experiment 2, even though the final reconstruction looks slightly worse than the one which was obtained with the data including all components of the wave field. Also the evolution of the internal parameter values inside the salt bodies follows a similar pattern and the final values are quite close to the correct reference values. This shows that the algorithm can be used even in cases where only pressure data are available, with only minor sacrifices in the quality of the final reconstruction.

Figures 11 and 12 show finally a situation where only pressure data is used, as in the previous case, but also additional Tikhonov regularization is added to the cost functional in addition to the Sobolev smoothing. We notice here again that the reconstruction looks overall much smoother compared to the one without the added Tikhonov term. However, the final result after 50 sweeps clearly has not recovered well some parts of the salt bodies. Our explanation of this observation is that some of the deeper parts of the salt domes in the reference profile have a very small impact on the (pressure-only) data least-squares misfit functional, which is comparable now in magnitude to the new impact of the Tikhonov term in the cost functional. Both are competing with each other, and obviously in the current simulation the minimization of the added Tikhonov term annihilates the deeper parts of the reconstructions of the salt bodies. Those regions usually have a very low sensitivity to the data, and the algorithm does not build those parts in the reconstructed profile since they would give rise to a significant increase in the added Tikhonov term.

5.4. Experiment 6: Level set reconstruction in a more complex situation

In this section we will demonstrate the performance of our level set based reconstruction approach in a slightly more complex situation. In order to test the stability and dependence on the model accuracy, we present a situation where the background model is intentionally more complex and inaccurately known. When using instead an inaccurate and simplified background model during the level set reconstruction process, the algorithm needs to cope with those complications. We will demonstrate that the level set approach is indeed able to handle such a situation and responds in a maybe surprising but very meaningful way.

The new and more complex model is displayed in Figure 13. In the left column of the figure the three profiles of $V_p$, $V_s$, and $\rho$ are shown together with the distribution of the sources and receivers (again 80 sources and 100 receivers). Notice that the background is more complex than in the previous model. A vertical cross-section through this background (not taking into account the embedded shapes) is displayed in the right column of the figure. The magenta profile with crosses represents the true reference background profile which is used for generating the data. Now, in contrary to our previous experiments, we do not have access to this true reference background profile during the shape reconstruction with the level set method. Instead we use a very rough linear approximation to it which is represented by the blue solid straight lines in the images on the right hand column. In particular, the two embedded layers of the reference background profile (one at depths between vertical pixel numbers 10-12 and the other one
Fig. 7: Numerical Experiment 3: Both, Sobolev and Tikhonov regularization are applied and all six components of the elastic wave field are measured at receiver locations. The first three rows show snapshots of the distribution of $V_p$ as well as the final level set function.
at pixel numbers 18-20) with high contrast to the background are assumed completely unknown to us. In addition, also the overall linear trend is not accurately represented by our background model. We will not attempt here to simultaneously reconstruct the background jointly with the level set representation of the shapes by an embedded pixel-based approach, even though that would be a plausible and doable approach in this situation. Instead, we apply our level set reconstruction scheme with the inaccurate background in order to verify whether the algorithm is still able to produce useful estimates of the shapes and contrast values, despite the errors and inaccuracies in the background model used during the reconstruction.

Figures 14 and 15 show the results of the level set reconstruction in this situation. No additional Tikhonov regularization is applied here during the reconstruction and the full data set is used. The initial guess consists of seven small discs fairly equidistantly distributed over the domain. It is chosen with the goal to fill the domain with sufficiently many seed objects which should reduce the overall evolution time without the need of any a priori knowledge about the final shape. Notice that the reconstructed shapes, displayed in Figure 14, are overall very close to the true reference shape. Interestingly, the inaccuracy in the background profile in the layers with vertical pixel numbers 10-12 are compensated by the algorithm in putting at this depth a thin layer of high contrast material of the shape material. This is of course plausible, since this is likely to reduce the overall data least-squares misfit. At depth 18-20, on the other hand, this is not feasible since the difference between background and true reference background values points into the other direction. The algorithm most likely will still attempt to compensate for this unknown layer. This might be the reason why in this numerical example the final internal parameter values of the shapes are not as accurate as they were in the previous test cases.

6. Summary and Conclusions

We have introduced in this paper a novel shape reconstruction approach combining level set regularization with a conservative elastic wave propagation model for calculating gradient directions. We have shown that this approach uses efficient time reversal concepts for calculating descent directions. It is different from the standard schemes for tackling the seismic FWI problem in various respects. Gradient directions make use of time derivatives instead of spatial derivatives of the wavefield at the given position. Moreover, it incorporates a narrow-band scheme which has the advantage that parameter or shape updates can be localized and are automatically blended out at regions which are not close to the current shape boundaries. Thereby, certain artefacts (e.g. close to the sources and receiver locations) can more easily be avoided. Certainly, some artefacts still remain, in particular small holes and imperfections inside the reconstructed salt bodies. Those most likely correspond to local minima. Even though they represent mainly cosmetic deficiencies which could in principle be removed by suitable image processing tools, they can also be
Fig. 9: Numerical Experiment 4: Only Sobolev regularization is applied and only pressure data are used.
avoided by adding an additional Tikhonov-Philips term to the least-squares data misfit functional. This has the effect of producing smoother level set functions representing the shapes, which eliminates most of those holes or imperfections. However, adding an additional term to the data misfit functional comes with a price to pay. When only pressure data are available, low sensitivity regions inside the earth might have only a low impact on the data misfit, which might be dominated by the impact they will have on the added Tikhonov term. This has the effect that objects located in low sensitivity regions tend to be concealed in the Tikhonov-Philips regularized reconstructions. Therefore, it might be better not to use such a Tikhonov-Philips term at all and completely rely on the Sobolev smoothing for stabilizing the reconstruction. Alternatively, a suitable balance between those two regularization approaches could be sought which provides an optimal trade-off between those competing effects or completely different Tikhonov terms could be tested instead.

In general, the results that we obtained with the combined shape and internal parameter reconstruction approach look very promising. We have observed that including the additional optimization of internal elastic parameters inside the salt bodies can help the algorithm getting out of certain local minima which might occur when only using a shape evolution with an incorrect choice of those internal elastic parameter values. In our experience, it seems to be advantageous to first concentrate on the shape evolution with fixed internal parameter values until the shape evolution reaches some form of stationary point. Then also the search for optimal internal elastic parameters should be incorporated into the optimization loop which not only can help the algorithm to escape some local minima, but also improves the reconstruction of the shapes. This is so since it does not seem to provide any advantage to start the search for optimal internal parameter values when the shape is still far away from some stationary point. This question however needs to be investigated further in order to arrive at some general recommendation.

We want to point out once more that this paper only provides a proof-of-concept study for the general approach addressed here. However, we are very confident that it can be generalized considerably to incorporate much more complex settings. It can be extended for example to incorporate also the simultaneous reconstruction of inhomogeneous elastic background profiles jointly with the shapes embedded into those profiles. Similar approaches have been demonstrated in different applications in [61][54][47].

Finally, regarding future work, it will be interesting to see the performance of the algorithm presented here in a more realistic 3D setting, ideally using real data. Also the incorporation of attenuation and anisotropy seems possible. A hybrid approach combining an acoustic wave propagation model in regions with small or vanishing shear modulus and a full elastic wave model in regions with significant shear wave contributions seems interesting and doable as well.

**References**

Fig. 11: Numerical Experiment 5: Both, Sobolev and Tikhonov regularization are applied and only pressure data are used.
Fig. 12: Numerical Experiment 5: Both, Sobolev and Tikhonov regularization are applied and only pressure data are used. The figure shows the evolution of the three internal parameter values as well as the least-squares cost functional.

Fig. 13: Numerical Experiment 6: This experiment uses a more complex situation. On the left, the three parameter distributions of the true reference model are displayed jointly with the source and receiver locations. On the right, vertical cross sections are shown through the reference model background parameters (excluding the shapes, magenta line with crosses) and through the background model parameters used throughout the level set inversion (blue solid line).
Fig. 14: Numerical Experiment 6: Only Sobolev regularization is applied to this more complex situation and all six components of the elastic wave field are measured at receiver locations. The inaccurate background model is used throughout the level set evolution.
Fig. 15: Numerical Experiment 6: Only Sobolev regularization is applied to this more complex situation and all six components of the elastic wave field are measured at receiver locations. The inaccurate background model is used throughout the level set evolution. The figure shows the evolution of the three internal parameter values as well as the least-squares cost functional.