On full seismic waveform inversion by descent methods in a lattice

Marcos Capistran, Miguel Angel Moreles, Joaquín Peña

CIMAT, Jalisco S/N Valenciana, Guanajuato, GTO 36240, MEXICO

Abstract: In this work we are concerned with the full waveform inversion problem. The problem is formulated as one of minimizing a nonlinear least squares functional. Assuming Fréchet differentiability we use the adjoint state approach to compute the gradient. To approximate local minima, we develop a discrete framework for descent methods in a finite difference lattice. We describe the methods of Gradient descent with line search and the positive definite secant update (BFGS) for computation in the lattice. To illustrate the methods numerical solutions of several examples in 1D are presented. In this case we carry out some analysis and provide a simple proof for identifiability of wave speeds using the spread and shrink argument. It is argued that we may build on this work and apply techniques such as regularization or bayesian inference in future investigations.

Keywords: Inverse Problems; Full Waveform Inversion; Identifiability; Descent method.

1. Introduction

Seismic inversion deals with the problem of determination of the structure and dynamics of Earth's interior by means of wave propagation in an elastic continuum. The underlying mathematical models are hyperbolic systems of partial differential equations obtained from Elastodynamics. In these models, the spatially varying coefficients of these PDEs describe the medium and hence its properties. Consequently, the inverse problem of concern is to estimate these coefficients from samples of quantities associated to the solution of the system of PDEs.

For instance, in oil industry applications reflection seismology is used to map petroleum deposits in the Earth's upper crust. The data are time series of pressure collected at points corresponding to locations of seismic sensors. A simplifying assumption is that the material does not support shear stress, thus restricting the problem to linear acoustics. In this context, we shall address the problem of estimating two spatially dependent parameters of the wave equation from measurements of the pressure, the time derivative of the solution, at a finite number of points in a given domain. We deal with full seismic waveform inversion, that is, we work with the fully nonlinear inverse problem.

The problem is by no means new but of great interest, a recent topical review is presented in Symes [15]. Even the 1D case is far from settled, see the numerical exploration for full seismic waveform inversion in Burstedde and Ghattas [3]. In line with this work we formulate the problem as as one of minimizing a nonlinear least squares functional. The minimization is carried out by descent gradient based methods.

We test the methods for several 1D examples, we show efficient identification of noisy as well as discontinuous parameters. It will become apparent that the algorithms are essentially dimension independent. Consequently, generalization to 2D and 3D scenarios is plausible.

There are numerous computational difficulties pertaining full waveform inversion, see Plassix [12] and Virieux [16] reviews, and the book by Fichtner [4]. Arguably, one of the most challenging difficulties of full waveform inversion is the fact that the associated least squares functional has spurious local minima. On the other hand, full waveform inversion requires efficient methods to compute both, the numerical solution of the wave equation and the gradient of the least squares functional defined by the misfit between the observed and the synthetic seismic data obtained by the numerical solution of the wave equation. The adjoint method is regarded as a standard tool to compute the

---

* Corresponding author: e-mail: marcos,moreles,joaquin@cimat.mx

gradient of the least squares functional, see [6,5,12,16]. Using variational calculus based on Gáteaux derivative, Sacks and Santosa [13] develop computational schemes for sound speed approximation in a similar 1D inverse problem. Existence and uniqueness conditions for the direct problem are given by Isakov [7]. Isakov also provides a theorem of existence and uniqueness for the inverse problem of determining the wave speed of the linear wave equation. Bamberger 1979 [1] has shown that the one dimensional inverse problem with a source function $f(z,t) = \delta(z)\delta(t)$ has a unique global solution provided exact data is used. Results on stability of the solution of the inverse problem are scarce at best, an exemption in 1D is the work of Symes [14].

Our contribution is in close relationship to these works, the outline is as follows. In Section 2 we present the formulation of the seismic inverse problem of interest and show the general setting for descent methods before discretization. In Section 3 we develop a finite difference framework to discretize the gradient and the Hessian of the least squares functional in 1D, 2D and 3D, we consider gradient descent with line search and the positive definite secant update for the Hessian, the so called BFGS method [10]. Numerical 1D examples of steepest descent and the BFGS methods are carried out in Section 4. In 1D the roles of time and space variables are interchangeable, thus we prove finite propagation speed in the space direction as the main step for identification in the spread and shrink argument of McLaughlin & Yoon [9]. This simple proof of a known result is the content of Section 5. We derive an expression for the gradient using the adjoint state approach in Section 6. We conclude our exposition with some conclusions and some thoughts on future research.

2. Problem Formulation

2.1. The direct problem.

Following Fichtner et al [5], we consider as a model for seismic energy propagation the scalar wave equation with variable coefficients

$$\rho(\mathbf{x})\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \nabla \cdot (\mu(\mathbf{x})\nabla u(\mathbf{x},t)) = g(\mathbf{x},t), \quad (1)$$
$$(\mathbf{x},t) \in \Omega \times (0,T).$$

Here Ω is a domain in $\mathbb{R}^3$. Initially the system is at rest, that is

$$u(\mathbf{x},0) = u_t(\mathbf{x},0) = 0, \qquad \mathbf{x} \in \Omega. \quad (2)$$

In principle, the domain Ω should be the unbounded half space $\{x_3 > 0\}$. For numerical simulation we shall consider a rectangular prism. Namely,

$$\Omega \ = (0,X) \times (0,Y) \times (0,Z) \subset \mathbb{R}^3,$$
$$Q_T = \Omega \times (0,T).$$

To make a well posed problem the following boundary conditions are customary

$$u(\mathbf{x},t) = 0, \ (\mathbf{x},t) \in \Gamma_1 \times [0,T]$$
$$\frac{\partial u(\mathbf{x},t)}{\partial n} = 0, \ (\mathbf{x},t) \in \Gamma_2 \times [0,T] \quad (3)$$

where $\Gamma_1 \cup \Gamma_2 = \Gamma$ is the boundary of Ω, and $\Gamma_1 \subset \{\mathbf{x} = (x,y,z) \in \mathbb{R}^3 : z = 0\}$.

2.2. The inverse problem.

Let us denote by $Y_s \times T_s, Y_s \subset \Gamma_1, T_s \subset [0,T]$, the source points, namely the support of $g(\mathbf{x},t)$. The data is gathered at the receiver points $Y_r \times T_r \subset \Gamma_1 \times [0,T]$ by means of a function d,

$$d \in L^2(Y_r \times T_r)$$

$$d \approx \frac{\partial u(\mathbf{x},t)}{\partial t} \equiv (\mathcal{M}u)(\mathbf{x},t), \qquad (\mathbf{x},t) \in Y_r \times T_r.$$

The problem of identification is to estimate ρ and μ from the data d. As remarked above, most approaches rely on some sort of linearization. Our aim is to consider the fully non linear problem in the least squares sense. Namely, we define the functional

$$J(\rho,\mu) = \frac{1}{2}\|\mathcal{M}u - d\|^2_{L^2(Y_r \times T_r)}$$

and consider the problem

$$\min J(\rho,\mu)$$
$$\text{subject to equations (1)-(3)} \quad (4)$$

The constrained optimization problem is solved by a descent method. Below is the general form of a continuous descent algorithm for approximation of a local minimum of J.

Algorithm. (Continuous descent)

Given a starting point (ρ_0,μ_0), a convergence tolerance ϵ, and $k \leftarrow 0$;
while $\|\nabla J((\rho_k,\mu_k))\| > \epsilon$;
    Compute search direction $p_k \equiv p_k(\nabla J((\rho_k,\mu_k)))$;
    Set $(\rho_{k+1},\mu_{k+1}) = (\rho_k,\mu_k) + \alpha_k p_k$;
    $k \leftarrow k + 1$;
end (while)

We consider gradient descent with line search and the positive definite secant update for the Hessian, the so called BFGS method.

In gradient descent

$$p_k = -\nabla J(\rho_k,\mu_k),$$

whereas for BFGS

$$p_k = -H_k \nabla J_k.$$

Here H_k is the well known inverse Hessian update.

Notice that in each iteration it is necessary to compute the gradient ∇J. The most efficient approach is by the adjoint state method. assuming Fréchet differentiability of the functional J we shall derive the following expression for the gradient,

$$\nabla J((\rho,\mu)) = \int_0^T \left(-\frac{\partial \lambda}{\partial t}\frac{\partial u}{\partial t}, \nabla \lambda \cdot \nabla u \right) dt.$$

The function λ is a Lagrange multiplier which is solution of the adjoint problem

$$\begin{aligned}
\rho(x)\frac{\partial^2 \lambda(x,t)}{\partial t^2} &- \nabla \cdot (\mu(x)\nabla\lambda(x,t)) \\
&= -\mathcal{M}^*\left(\mathcal{M}u - d\right), \quad (x,t) \in \Omega_T. \\
\lambda(x,T) &= \lambda_t(x,T) = 0, \qquad x \in \Omega. \\
\lambda(x,t) &= 0, \qquad (x,t) \in \Gamma_1 \times [0,T]. \\
\frac{\partial \lambda(x,t)}{\partial n} &= 0, \qquad (x,t) \in \Gamma_2 \times [0,T].
\end{aligned} \quad (5)$$

Computation of the gradient depends upon the solutions of wave equations for u and λ. Consequently, we shall develop a framework for discrete versions of the continuous descent methods suitable for fast computation.

3. Descent methods in a lattice

Some readers may find this section somewhat elementary, but besides clarity, it presents in some detail the discrete version of the method of solution. Numerical implementations, serial and parallel, are readily applied.

3.1. Inner product spaces in lattices

Consider the uniform grid

$$\begin{aligned}
0 &= x_0 < x_1 < \ldots < x_{M=1} = X, \\
0 &= y_0 < y_1 < \ldots < y_{N+1} = Y, \\
0 &= z_0 < z_1 < \ldots < z_{K+1} = Z, \\
0 &= t_0 < t_1 < \ldots < t_{L+1} = T.
\end{aligned}$$

Let us define the vector space

$$V = span\left\{ E_{\mathbf{m}}^l \right\},$$

where $0 \le m_1, m_2, m_3, l \le M+1, N+1, K+1, L+1$, and $\mathbf{m} = (m_1, m_2, m_3)$. Let us denote

$$\sum_{\mathbf{m}} = \sum_{m_1=0}^{M+1} \sum_{m_2=0}^{N+1} \sum_{m_3=0}^{K+1}.$$

For $u \in V$,

$$u = \sum_{\mathbf{m}} \sum_{l=0}^{L+1} u_{\mathbf{m}}^l E_{\mathbf{m}}^l.$$

Recall

$$(u,v)_{L^2(Q_T)} = \int_\Omega \int_{(0,T)} u(\mathbf{x},t)v(\mathbf{x},t)d\mathbf{x}dt.$$

From a quadrature on the grid we equip V with the inner product

$$(u,v)_{\Delta V \, \Delta t} = \Delta V \, \Delta t \sum_{\mathbf{m}} \sum_{l=0}^{L} u_{\mathbf{m}}^l v_{\mathbf{m}}^l,$$

where $\Delta V = \Delta x \Delta y \Delta z$.

For space variables we consider the vector space

$$V_\Omega = \text{span}\left\{ E_{(m_1,m_2,m_3)} : 0 \le m_1 \le M+1, \\ 0 \le m_2 \le N+1, 0 \le m_3 \le K+1 \right\}$$

and argue as before. We discretize the Initial Boundary Value Problem (IBVP) (1)-(3) using finite differences and consider the corresponding inverse problem in the inner product space $(V, (\cdot,\cdot)_{\Delta V \, \Delta t})$, the space-time lattice.

Remark. Let H be a Hilbert space and $A : H \to H$ a linear operator. We have that

$$(Au,v)_H = (u, A^*v)_H \qquad (6)$$

where A^* is the adjoint of A. We shall use this expression repeatedly.

3.2. The source term in the adjoint equation

The discrete version of the IBVP (1)-(3) is straightforward. For the adjoint equation let us derive an expression for the source term $\mathcal{M}^*\left(\mathcal{M}u - d\right)$ in the space-time lattice.

3.2.1. The discrete $\mathcal{M}$ operator

Recall that

$$\left(\mathcal{M}u\right)(\mathbf{x},t) = \frac{\partial u(\mathbf{x},t)}{\partial t}, \quad (\mathbf{x},t) \in Y_r \times T_r \subset \Gamma_1 \times [0,T].$$

Let us consider a receiver point $\mathbf{x} \in \Gamma_1$, thus $\mathbf{x} = (x_1, x_2, 0)$. Let us define the *trace operator* $\tau_{(\cdot,\cdot,0)} : V \to V$ by

$$\left(\tau_{(\cdot,\cdot,0)}\right)u = \sum_{m_1=0}^{M+1} \sum_{m_2=0}^{N+1} \sum_{l=0}^{L+1} u_{(m_1,m_2,0)}^l E_{(m_1,m_2,0)}^l$$

We also require a discrete version of the time derivative. Using an approximation backwards in time we have

$$\frac{\partial}{\partial t} : V \to V,$$

$$\frac{\partial}{\partial t}u = \sum_{\mathbf{m}} \sum_{l=1}^{L+1} \frac{u_{\mathbf{m}}^l - u_{\mathbf{m}}^{l-1}}{\Delta t} E_{\mathbf{m}}^l.$$

Note the zero components

$$\left(\frac{\partial}{\partial t}u\right)^0_{\mathbf{m}} = 0.$$

Let $I^0 : V \to V, \quad S^- : V \to V$ given by

$$I^0 u = \sum_{\mathbf{m}} \sum_{l=1}^{L+1} u^l_{\mathbf{m}} E^l_{(m_1, m_2, m_3)}$$

and

$$S^- u = \sum_{\mathbf{m}} \sum_{l=1}^{L+1} u^{l-1}_{\mathbf{m}} E^l_{\mathbf{m}}.$$

A simple decomposition of the discrete time derivative is as follows

$$\frac{\partial}{\partial t} = \frac{1}{\Delta t}\left(I^0 - S^-\right)$$

We conclude that the discrete $\mathcal{M}$ operator is the composition

$$\mathcal{M} = \tau_{(\cdot,\cdot,0)} \circ \frac{\partial}{\partial t} \qquad (7)$$

3.2.2. The discrete $\mathcal{M}^*$ operator

From (7) it follows that

$$\mathcal{M}^* = \frac{1}{\Delta t}\left(\left(I^0\right)^* - \left(S^-\right)^*\right) \circ \left(\tau_{(\cdot,\cdot,0)}\right)^*.$$

Consider the operator $S^+ : V \to V$ given by

$$S^+ v = \sum_{\mathbf{m}} \sum_{l=0}^{L} v^{l+1}_{(m_1, m_2, m_3)} E^l_{(m_1, m_2, m_3)}$$

By (6) it is readily seen that

$$\left(S^-\right)^* = S^+.$$

Also

$$\left(I^0_{(\cdot)}\right)^* = \left(I^0_{(\cdot)}\right),$$

and

$$\left(\tau^{(\cdot)}_{(\cdot,\cdot,0)}\right)^* \equiv \tau^{(\cdot)}_{(\cdot,\cdot,0)}.$$

Consequently

$$\mathcal{M}^* = \left(\frac{\partial}{\partial t}\right)^* \circ \left(\tau^{(\cdot)}_{(\cdot,\cdot,0)}\right)^*$$

or

$$\mathcal{M}^* = \frac{1}{\Delta t}\left(I^0_{(\cdot)} - S^+\right) \circ \tau^{(\cdot)}_{(\cdot,\cdot,0)}.$$

3.3. The discrete gradient ∇J

The gradient ∇J is a function in space variables. Thus we consider the space lattice $V_\Omega = span\{E_{(m_1, m_2, m_3)}\}$, where $0 \le m_1, m_2, m_3 \le M+1, N+1, K+1$. Let $u, v \in V_\Omega$. Then

$$u = \sum_{\mathbf{m}} \sum_{l=0}^{L+1} u^l_{\mathbf{m}} E^l_{\mathbf{m}}, \qquad v = \sum_{\mathbf{m}} \sum_{l=0}^{L+1} v^l_{\mathbf{m}} E^l_{\mathbf{m}}.$$

We define the (pointwise) product by the expression

$$uv = \sum_{\mathbf{m}} \sum_{l=0}^{L+1} \left(u^l_{\mathbf{m}} v^l_{\mathbf{m}}\right) E^l_{\mathbf{m}}.$$

The discrete integral operator

$$\int_0^T : V \to V_\Omega$$

is

$$\left(\int_0^T u\right)_{\mathbf{m}} = \left(\Delta t \sum_{l=0}^{L} u^l_{\mathbf{m}}\right) E_{\mathbf{m}}.$$

To construct partial derivatives we define the derivative $D_{j,\mathbf{m}} : V_\Omega \to V_\Omega$ in the jth direction in each $\mathbf{m}$ node as follows. For $1 \le m_1 \le M$,

$$D_{(1,\mathbf{m})}u = \frac{u_{(m_1+1, m_2, m_3)} - u_{(m_1-1, m_2, m_3)}}{2\Delta x} E_{\mathbf{m}}.$$

For $m_1 = 0$,

$$D_{(1,\mathbf{m})}u = \frac{4u_{(1, m_2, m_3)} - 3u_{(0, m_2, m_3)} - u_{(2, m_2, m_3)}}{2\Delta x} E_{\mathbf{m}}.$$

This is the second order approximation for the derivative in the boundary. Similarly $m_2, m_3 = 0$.

We obtain jth derivatives in the space lattice

$$D_j = \sum_{\mathbf{m}} D_{(j,\mathbf{m})}, \qquad j = 1, 2, 3.$$

and the discrete gradient $\nabla : V_\Omega \to (V_\Omega)^3$,

$$\nabla = (D_1, D_2, D_3).$$

Recall the continuous gradient

$$\nabla J((\rho, \mu)) = \left(\int_0^T -\frac{\partial \lambda}{\partial t}\frac{\partial u}{\partial t} dt, \int_0^T \nabla \lambda \cdot \nabla u \, dt\right).$$

The discrete approximation is for each component

$$-\left(\int_0^T \frac{\partial \lambda}{\partial t}\frac{\partial u}{\partial t}\right)_{\mathbf{m}} = -\left(\frac{1}{\Delta t}\sum_{l=0}^{L}\left[\left(I^0_{(\cdot)} - S^-\right)\lambda\right]^l_{\mathbf{m}}\right.$$
$$\left.\left[\left(I^0_{(\cdot)} - S^-\right)u\right]^l_{\mathbf{m}}\right) E_{\mathbf{m}}$$

and

$$\left(\int_0^T \nabla \lambda \cdot \nabla u\right)_{\mathbf{m}} = \left(\int_0^T \sum_{k=1}^{3} (D_k \lambda)(D_k u)\right)_{\mathbf{m}}$$

3.4. Finite difference inverse Hessian approximation

Let us derive an expression for matrix vector multiplication to be applied in the BFGS method.

If $H \in L\left(V_\Omega \times V_\Omega, V_\Omega \times V_\Omega\right)$ the matrix's components are given by

$$H^{(\mathbf{a},\mathbf{b})}_{(\mathbf{m},\mathbf{n})} = \left(\left(E_\mathbf{a}, E_\mathbf{b}\right), H\left(E_\mathbf{m}, E_\mathbf{n}\right)\right)_{\Delta V}.$$

This yields the following expression for the required product

$$H(p,q) = \sum_\mathbf{a} \sum_\mathbf{b} \left(\sum_\mathbf{m} \sum_\mathbf{n} (p,q)_{(\mathbf{m},\mathbf{n})} H^{(\mathbf{a},\mathbf{b})}_{(\mathbf{m},\mathbf{n})}\right) (E_\mathbf{a}, E_\mathbf{b}).$$

Remark. We have developed a discrete scheme which is atomic and highly parallelizable. For instance, the approach in Ortigosa et al [11] seems plausible.

4. Numerical results in 1D

In this section we illustrate that the discrete framework developed above can be applied effectively for parameter identification. Here we content ourselves with applications in 1D for synthetic data.
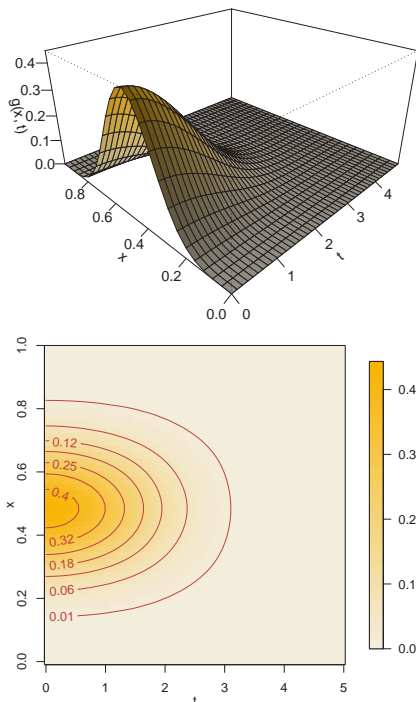

Figure 1 Function $g(x,t)$ with compact support.

We have run extensive experiments with great success. From our sample we have chosen examples with smooth, noisy and discontinuous coefficients. For computation we have used a PC with an AMD Athlon Dual Core processor running at 1 MHz.

Our initial motivation was to evaluate the performance of BFGS in this problem of identification. The results are promising for extension to several dimensions. For comparison we also run gradient descent with line search.

For the numerical tests the spatial domain is $\Omega = [0,1]$ and the time domain is $[0,5]$. Fig. 1 shows the function $g(x,t)$ which is a product of two cubic b-splines, $g(x,t) = \beta(2x-1)\beta(t/2)$, where

$$\beta(x) = \begin{cases} \frac{1}{6}(3|x| - 6x^2 + 4) & 0 \leq |x| < 1, \\ \frac{1}{6}(-|x|^3 + 6x^2 - 12|x| + 8) & 1 \leq |x| < 2, \\ 0 & \text{otherwise.} \end{cases}$$

The position of the receiver is $x_r = 0.1$. At each time step t_l, $l = 0, ..., L+1$, a measurement is recorded. The problem (1)-(3) is solved using a Crank-Nicholson scheme. One can assume that near $x = 0$ the values of ρ and μ are known. From this information, one can use linear extrapolation to set the initial estimation of ρ and μ.

### 4.1. Reconstruction of ρ with known constant μ.

From top to bottom, Fig. 2 shows the data d, the initial approximation obtained using the first estimation $\rho^{(0)}$, and the final approximation. The second plot shows the target function ρ, the initial guess $\hat{\rho}^{(0)}$ and the solution obtained with the proposal method using the BFGS algorithm. Although the initialization near $x = 1$ is not suitable, the method is able to correct these values and obtain a good approximation of the true values of ρ. The last two plot in Fig. 2 shows the behavior of the function $J(\hat{\rho}^{(i)}, \mu)$ and the reconstruction error $\|\hat{\rho}^{(i)} - \rho\|/\sqrt{M}$ evaluated in the ith estimated function $\hat{\rho}^{(i)}$ using BFGS and gradient descent. According to these plots, the results are similar, but the BFGS requires fewer iterations.

Table 1 Comparison of the reconstruction error and the elapsed time using the two optimization strategies.

Example	Algorithm	$\|\rho - \hat{\rho}\|/\sqrt{M}$	Time
1	GD	0.00044	3.63 minutes
(Fig. 2)	BFGS	0.00038	0.64 minutes
2	GD	0.00058	7.26 minutes
(Fig. 3)	BFGS	0.00070	1.52 minutes
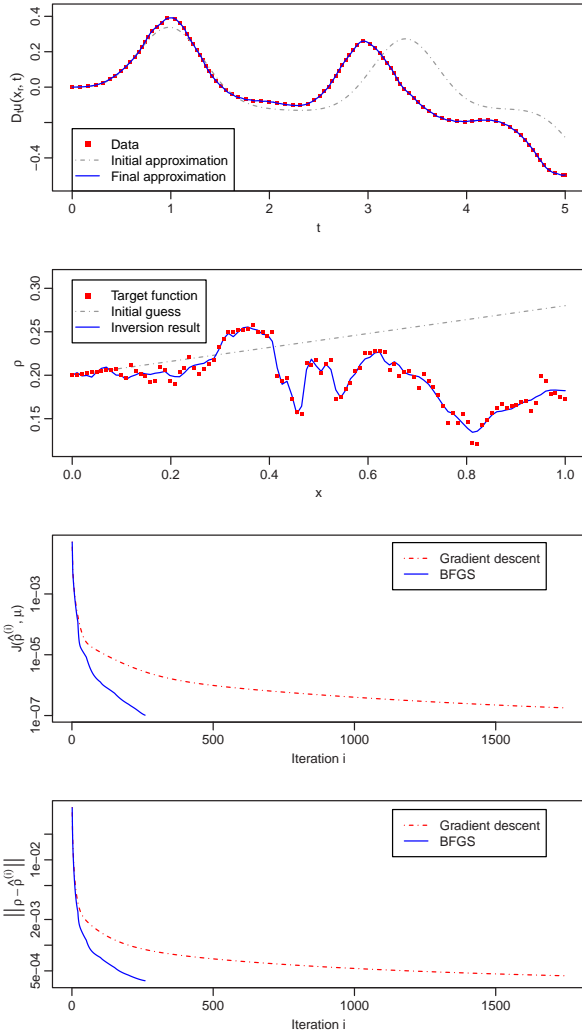
Figure 2 Reconstruction of non smooth data.
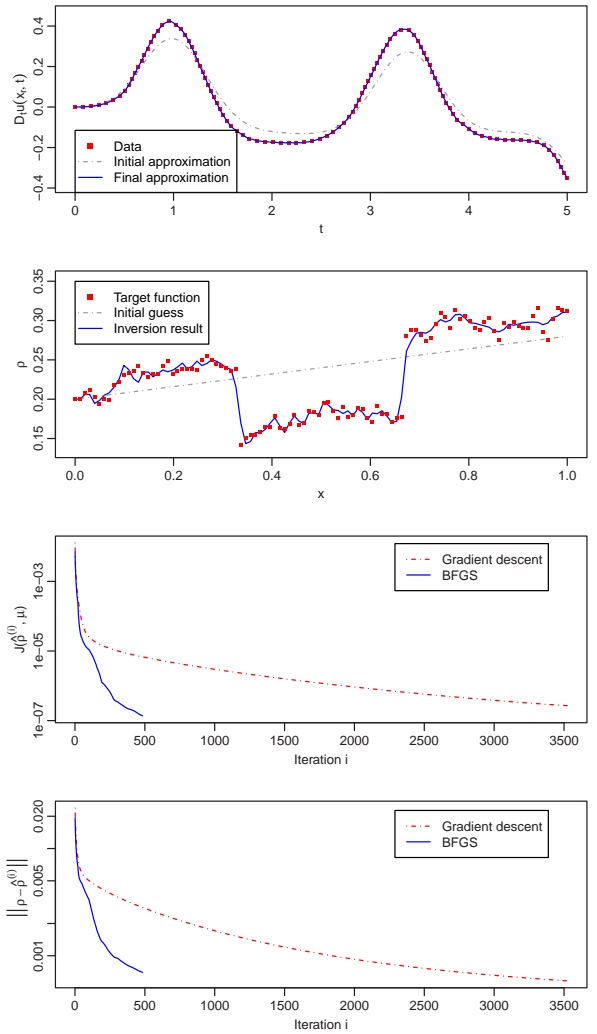
Figure 3 Reconstruction of discontinuous spatial data.

Fig. 3 shows the results when the function ρ has discontinuities. The method is able to detect the points where breaks occur. The performance of both optimization strategies is summarized in Table 1 for the two experiments.

Remark. At present we are not able to show convergence analysis of the proposed method neither stability. As an alternative we have carried out numerical experiments on stability. As example of this experiments see Fig. 4 for a comparison of the error when the problem is solved by BFGS using three different discretizations. The number of iterations is similar but the computation time in minutes were 1.52, 6.78, 42.05 from low to high resolution, respectively.

4.2. Reconstruction of μ with known constant ρ.

In Fig. 5 is showed the reconstruction of the function μ when ρ is a known constant. In general, the reconstructed function $\hat{\mu}$ is smoother than the true function μ. The performance of the two optimization algorithms for this experiment is showed in Table 2.

Table 2 Comparison of the reconstruction error and the elapsed time using the two optimization strategies.

Algorithm	$\|\mu - \hat{\mu}\|/\sqrt{M}$	Time
GD	0.00196	11.11 minutes
BFGS	0.00081	3.16 minutes
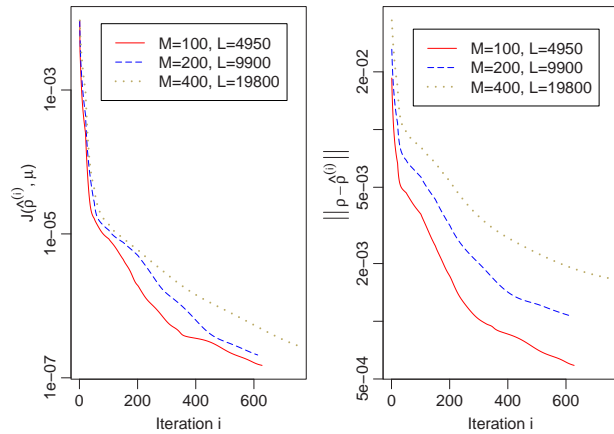
Figure 4 Behavior of the error in each iteration using BFGS in three different discretizations.

4.3. Reconstruction of ρ and μ.

When ρ and μ are unknown, the reconstructed functions $\hat{\mu}$ and $\hat{\rho}$ are less accurate, as shown in Figure 6, although these functions satisfy the condition of reproducing the data (Figure 7). Also the convergence to the solution is slower. Using BFGS the elapsed time was 5.48 minutes. In the end, the data d are reproduced with a precision that is similar to previous cases. In general, this is the only indicator we have to accept the solution.

5. Identifiability in 1D

We assume that

$$\frac{\partial u(0,t)}{\partial t}, \quad t \in [0, T),$$

is known. Since u is initially at rest, so is

$$u(0,t), \quad t \in [0, T).$$

We shall show that with this data and constant μ density is identifiable. The result is not new, but a simple proof can be given following the spread argument of McLaughlin & Yoon [9]. The argument consists on applying in tandem a result on finite propagation speed and a unique continuation principle. We do not aim for generality, thus to simplify the proofs we shall assume appropriate smoothness.

5.1. Finite speed of propagation

For our problem of identification we need to consider x as the time variable in order to establish a result on finite speed of propagation. See O. A. Ladyzhenskaya [8].
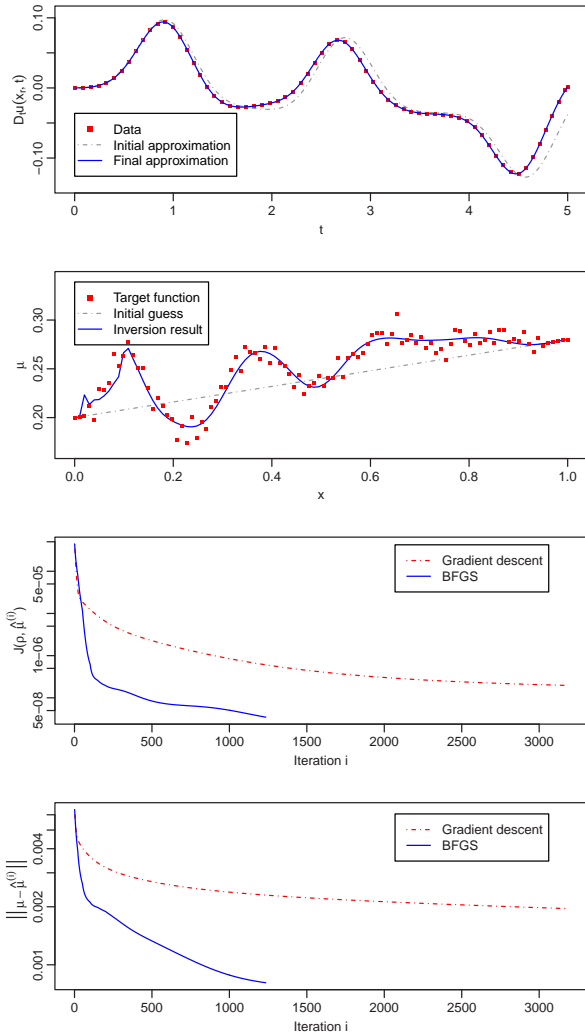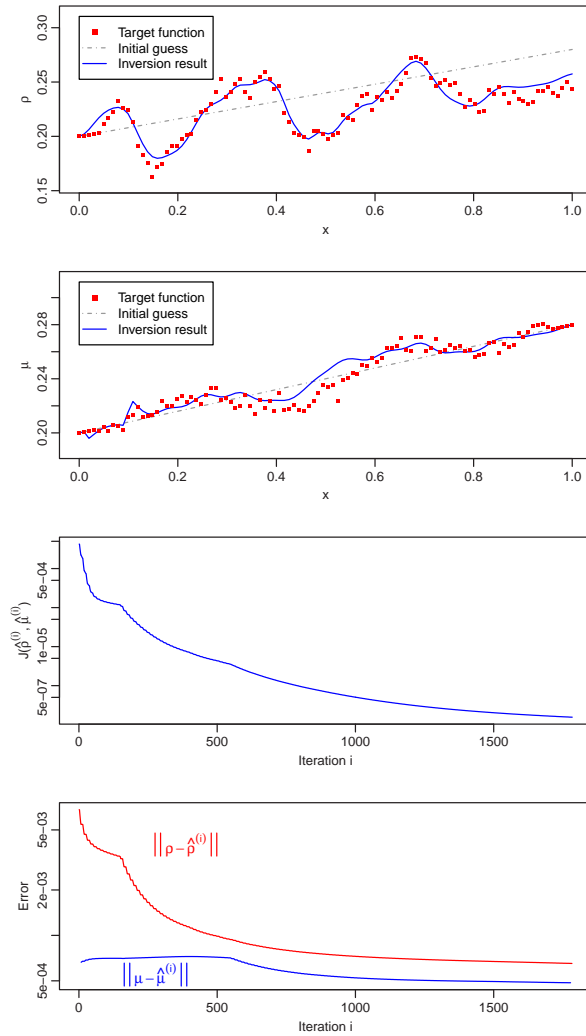

Figure 5 Reconstruction of a non smooth function μ.

Theorem 1.*Assume that $\rho \in C^1([0, X])$ and $\mu \in C^1([0, X])$ satisfy $\rho(x), \mu(x) \geq \alpha_0 > 0$. Let $u \in H^2((0, X) \times (0, T))$ be a solution to the wave equation*

$$(\mu u_x)_x - \rho u_{tt} = f, \quad in \ (0, X) \times (0, T).$$

Then for any open interval $B_\varepsilon(t_0) = (t_0 - \varepsilon, t_0 + \varepsilon) \subset (0, X)$, u has a finite speed of propagation in $(0, X) \times B_\varepsilon(t_0)$ with the maximum speed

$$v = \max_{x \in [0, X]} \sqrt{\frac{\rho(x)}{\mu(x)}}$$

Figure 7 Data with reconstructed μ and ρ using BFGS.

$$\frac{\partial}{\partial x}\left(\mu u_x\right)^2 - 2\frac{\partial}{\partial t}\left(\rho\mu u_x u_t\right) + \rho\mu\frac{\partial}{\partial x}\left(u_t\right)^2 = 2\mu u_x f.$$

$$\frac{\partial}{\partial x}\left(\mu u_x\right)^2 - 2\frac{\partial}{\partial t}\left(\rho\mu u_x u_t\right) + \rho\mu\frac{\partial}{\partial x}\left(u_t\right)^2 +$$
$$\left(u_t\right)^2\frac{\partial}{\partial x}\left(\rho\mu\right) - \left(u_t\right)^2\frac{\partial}{\partial x}\left(\rho\mu\right) = 2\mu u_x f.$$

We obtain the expression

$$\frac{\partial}{\partial x}\left(\mu u_x\right)^2 + \frac{\partial}{\partial x}\left(\rho\mu\left(u_t\right)^2\right) - 2\frac{\partial}{\partial t}\left(\rho\mu u_x u_t\right)$$
$$- \left(u_t\right)^2\frac{\partial}{\partial x}\left(\rho\mu\right) = 2\mu u_x f,$$

which can be written in the form

$$\nabla\cdot\left(\left(\mu u_x\right)^2 + \rho\mu\left(u_t\right)^2, -2\rho\mu u_x u_t\right)$$
$$= \left(u_t\right)^2\frac{\partial}{\partial x}\left(\rho\mu\right) + 2\mu u_x f.$$

Integrate over Λ_s to obtain

$$\int_{\Lambda_s}\nabla\cdot\left(\left(\mu u_x\right)^2 + \rho\mu\left(u_t\right)^2, -2\rho\mu u_x u_t\right)$$
$$= \int_{\Lambda_s}\left[\left(u_t\right)^2\frac{\partial}{\partial x}\left(\rho\mu\right) + 2\mu u_x f\right].$$

The set Λ_s is a trapezium with two sides parallel to the t-axis, C_0 and C_s respectively. We shall denote by Γ_1 the side C_0 and number the other sides counterclockwise. Then, by the divergence theorem

$$-\int_{\Gamma_1}\left(\left(\mu u_x\right)^2 + \rho\mu\left(u_t\right)^2\right) + \int_{\Gamma_r}\left(\left(\mu u_x\right)^2 + \rho\mu\left(u_t\right)^2\right)$$
$$+\int_{\Gamma_2}\frac{1}{\|(v,-1)\|}\left[\left(\left(\mu u_x\right)^2 + \rho\mu\left(u_t\right)^2\right)v + 2\rho\mu u_x u_t\right]$$
$$+\int_{\Gamma_4}\frac{1}{\|(v,1)\|}\left[\left(\left(\mu u_x\right)^2 + \rho\mu\left(u_t\right)^2\right)v - 2\rho\mu u_x u_t\right]$$
$$= \int_{\Lambda_s}\left[\left(u_t\right)^2\frac{\partial}{\partial x}\left(\rho\mu\right) + 2\mu u_x f\right].$$

Figure 6 Simultaneous reconstruction of μ and ρ using BFGS.

Proof. Since $\mu(x)\geq\alpha_0 > 0$, we may regard x as the time variable. For $s < \frac{\varepsilon}{v}$, we define

$$C_s = \left(t_0 - \left(\varepsilon - vs\right), t_0 + \left(\varepsilon - vs\right)\right) \times \{x = x_0 + s\},$$
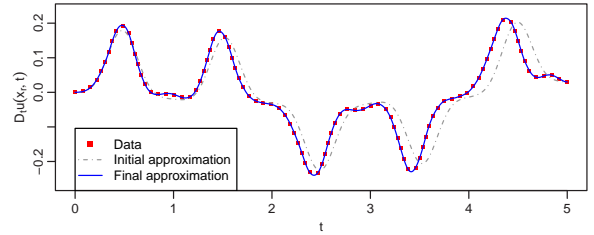$$\Lambda_s = \bigcup_{0 < \xi < s} C_\xi.$$

Consider the equation

$$\left(\mu u_x\right)_x - \rho u_{tt} = f.$$

Multiply it by $2\mu u_x$,

$$2\mu u_x\left(\mu u_x\right)_x - 2\rho\mu u_x u_{tt} = 2\mu u_x f.$$

$$2\mu u_x\left(\mu u_x\right)_x - 2\rho\mu u_x u_{tt} - 2\rho\mu u_{xt}u_t + 2\rho\mu u_{xt}u_t = 2\mu u_x f.$$

$$-\int_{\Gamma_1} \left((\mu u_x)^2 + \rho\mu (u_t)^2\right) + \int_{\Gamma_3} \left((\mu u_x)^2 + \rho\mu (u_t)^2\right)$$

$$+ \int_{\Gamma_2} \frac{v}{\|(v,-1)\|} \left[\left((\mu u_x)^2 + \rho\mu (u_t)^2\right) + \frac{2}{v}\rho\mu u_x u_t\right]$$

$$+ \int_{\Gamma_4} \frac{v}{\|(v,1)\|} \left[\left((\mu u_x)^2 + \rho\mu (u_t)^2\right) - \frac{2}{v}\rho\mu u_x u_t\right]$$

$$= \int_{\Lambda_s} \left[(u_t)^2 \frac{\partial}{\partial x}(\rho\mu) + 2\mu u_x f\right].$$

$$(8)$$

Consider the inequality

$$\left|\frac{2}{v}\rho\mu u_x u_t\right| = 2 |\mu u_x| \left|\frac{\rho}{v}u_t\right| \leq (\mu u_x)^2 + \left(\frac{\rho}{v}\right)^2 (u_t)^2.$$

Hence

$$\left(\frac{\rho}{v}\right)^2 \leq \rho\mu$$

if and only if

$$\sqrt{\frac{\rho}{\mu}} \leq v.$$

From (8) we obtain

$$\int_{\Gamma_3} \left((\mu u_x)^2 + \rho\mu (u_t)^2\right) - \int_{\Gamma_1} \left((\mu u_x)^2 + \rho\mu (u_t)^2\right)$$

$$\leq \int_{\Lambda_s} \left[(u_t)^2 \frac{\partial}{\partial x}(\rho\mu) + 2\mu u_x f\right]. \quad (9)$$

But ρ and μ are C^1 functions, so

$$\int_{\Lambda_s} \left[(u_t)^2 \frac{\partial}{\partial x}(\rho\mu) + 2\mu u_x f\right]$$

$$\leq c \int_{\Lambda_s} \left[(u_t)^2 + (\mu u_x)^2 + |f|^2\right]$$

Since ρ and μ are bounded from below, it follows from (9) that

$$\int_{\Gamma_3} \left[(u_x)^2 + (u_t)^2\right] \leq c \int_{\Gamma_1} \left((u_x)^2 + (u_t)^2\right)$$

$$+ c \int_{\Lambda_s} |f|^2 + c' \int_{\Lambda_s} \left[(u_t)^2 + (u_x)^2\right]. \quad (10)$$

Let us define the *energy*

$$E(\sigma) = \int_{\Gamma_\sigma} \left((u_x)^2 + (u_t)^2\right).$$

We can write (10) in the form

$$E(s) \leq c \left[E(0) + \int_{\Lambda_s} |f|^2\right] + c' \int_0^s E(\sigma)d\sigma.$$

Thus, if $f \equiv 0$ in Λ_s, by Gronwall's lemma

$$E(s) \leq cE(0) \exp(c's).$$

We conclude finite speed v of propagation.

5.2. Uniqueness of wave speeds

Definition. Let Ω be an open domain in $\mathbb{R}^n$, we say that $u \in H^1_{loc}(\Omega)$ has a unique continuation principle in Ω if $u = 0$ in an open subset of Ω implies that $u = 0$ in Ω.

It is well known that elliptic equations have a unique continuation principle. We shall apply this principle to the one dimensional equation $u_{tt} = 0$, where t is regarded as space variable.

Theorem 2.. *Let μ be a known positive constant. Assume that $\rho_j \in C^1([0,X])$, $j = 1,2$ satisfy $\rho_j(x) \geq \alpha_0 > 0$. Let $u \in H^2((0,X) \times (0,T))$ be a common solution to the wave equation*

$$\mu u_{xx} - \rho_j u_{tt} = f, \quad in \ (0,X) \times (0,T).$$

with zero Cauchy data

$$u(x,0) = u_t(x,0) = 0, \ x \in (0,X).$$

Let $t_0 > 0$ and $\varepsilon > 0$ be given. Consider the boundary data condition

$$u(0,t) = 0, \ t \in (t_0 - \varepsilon, t_0 + \varepsilon).$$

Then we have

$$\rho_1 = \rho_2, \qquad in \ (0,X) \backslash I_E,$$

where

$$I_E = \bigcup \left\{ I \subset (0,X) : \ I \ open, \ \|u\|_{L^2(I \times (0,T))} = 0 \right\}.$$

Proof. Let $(0,X)$ be expressed by the union of disjoint subsets $I = I^0 \cup I^+ \cup I^-$, where

$$I^0 = \{x \in (0,X) : \rho_1(x) = \rho_2(x)\}.$$

$$I^\pm = \{x \in (0,X) : \rho_1(x) \gtrless \rho_2(x)\}.$$

We will show that $I^+ \cup I^- \subset I_E$. Fix any point $x_0 \in I^+$. Since $\rho_1 - \rho_2 \in C^0([0,X])$, and I^+ is an open subset of $(0,X)$, there exists an open interval $I_\varepsilon(x_0) \subset I^+$ on which we have

$$\alpha_1 < \rho_1 - \rho_2 < \alpha_2 \quad \text{for some } \alpha_1, \alpha_2 > 0.$$

We assume that $u(x,t)$ is a common solution of the wave equation for $\rho_j(x)$, $j = 1,2$.

$$\mu u_{xx} - \rho_1 u_{tt} = f,$$

$$\mu u_{xx} - \rho_2 u_{tt} = f.$$

Subtracting

$$(\rho_2 - \rho_1) u_{tt}(x,t) = 0.$$

Thus, if $x \in I_\varepsilon(x_0)$ u satisfies the equation

$$u_{tt}(x,t) = 0, \quad t \in (0,T). \quad (11)$$

But because of finite propagation speed, $u \equiv 0$ in

$$C_x = \{x\} \times (t_0 - (\varepsilon - vs), t_0 + (\varepsilon - vs)),$$

where

$$v = \max_{y \in [0,X]} \sqrt{\frac{\rho_1(y)}{\mu}}$$

If u is solution of (11), then it satisfies a unique continuation principle and $u \equiv 0$ in $I_\varepsilon(x_0) \times (0,T)$. Consequently $I^+ \subset I_E$. Similarly we have $I^- \subset I_E$, implying that $(0,X) \backslash I_E \subset (0,X) \backslash (I^+ \cup I^-) \subset I^0$, which completes the proof.

6. Gradient by the adjoint state method

By assuming Fréchet differentiability, we derive an expression for the gradient using the adjoint state approach. Our contention is that Fréchet differentiability adds clarity and generality to gradient computation instead of that of Gáteaux.

For the constrained minimization problem (4) we consider the Lagrangian

$$\mathcal{L}(\rho, \mu, u; \lambda) = \frac{1}{2} \|\mathcal{M}u - d\|^2_{L^2(Y_r \times T_r)}$$
$$+ \left\langle \rho \frac{\partial^2 u}{\partial t^2} - \nabla \cdot (\mu \nabla u) - g, \lambda \right\rangle_{L^2(\Omega_T)},$$

where λ is a Lagrange multiplier.

Let us assume that $\mathcal{L}(\rho, \mu, u; \lambda)$ is Fréchet differentiable. We denote by $D_j \mathcal{L}$ the partial derivative of $\mathcal{L}$ with respect to the j variable.

We shall use freely the rules of differentiation in the sense of Fréchet. See Dieudonné [2].

$$D\mathcal{L}(\rho, \mu, u; \lambda)(\xi, \varsigma, \eta) = D_1\mathcal{L}(\rho, \mu, u; \lambda)\xi$$
$$+ D_2\mathcal{L}(\rho, \mu, u; \lambda)\varsigma + D_3\mathcal{L}(\rho, \mu, u; \lambda)\eta,$$

where

$$D_1\mathcal{L}(\rho, \mu, u; \lambda)\xi = \left\langle \xi \frac{\partial^2 u}{\partial t^2}, \lambda \right\rangle_{L^2(\Omega_T)},$$

$$D_2\mathcal{L}(\rho, \mu, u; \lambda)\varsigma = \left\langle -\nabla \cdot (\varsigma \nabla u), \lambda \right\rangle_{L^2(\Omega_T)},$$

$$D_3\mathcal{L}(\rho, \mu, u; \lambda)\eta = \left\langle \mathcal{M}\eta, \mathcal{M}u - d \right\rangle_{L^2(Y_r \times T_r)} +$$
$$\left\langle \rho \frac{\partial^2 \eta}{\partial t^2} - \nabla \cdot (\mu \nabla \eta), \lambda \right\rangle_{L^2(\Omega_T)}.$$

Given ρ, μ let $u(\rho, \mu)$ be the solution of the IBVP and define

$$U(\rho, \mu) = (\rho, \mu, u(\rho, \mu)).$$

Thus

$$J(\rho, \mu) = \mathcal{L}(U(\rho, \mu); \lambda).$$

By the chain rule

$$DJ(\rho, \mu)(\xi, \varsigma) = D\mathcal{L}(U(\rho, \mu); \lambda)DU(\rho, \mu)(\xi, \varsigma),$$

but

$$DU(\rho, \mu)(\xi, \varsigma) = D_1U(\rho, \mu)\xi + D_2U(\rho, \mu)\varsigma,$$

$$D_1U(\rho, \mu)\xi = (\xi, 0, D_1u(\rho, \mu)\xi),$$

$$D_2U(\rho, \mu)\varsigma = (0, \varsigma, D_2u(\rho, \mu)\xi\zeta).$$

Define $v = Du(\rho, \mu)(\xi, \varsigma)$. From

$$DJ(\rho, \mu)(\xi, \varsigma) = D\mathcal{L}(U(\rho, \mu); \lambda)(\xi, \varsigma, Du(\rho, \mu)(\xi, \varsigma))$$

$$\equiv D\mathcal{L}(U(\rho, \mu); \lambda)(\xi, \varsigma, v)$$

we write

$$DJ(\rho, \mu)(\xi, \varsigma) = \left\langle \xi \frac{\partial^2 u}{\partial t^2}, \lambda \right\rangle_{L^2(\Omega_T)}$$
$$+ \left\langle -\nabla \cdot (\varsigma \nabla u), \lambda \right\rangle_{L^2(\Omega_T)}$$
$$+ \left\langle \mathcal{M}v, \mathcal{M}u - d \right\rangle_{L^2(Y_r \times T_r)}$$
$$+ \left\langle \rho \frac{\partial^2 v}{\partial t^2} - \nabla \cdot (\mu \nabla v), \lambda \right\rangle_{L^2(\Omega_T)}$$
$$= \left\langle \xi \frac{\partial^2 u}{\partial t^2}, \lambda \right\rangle_{L^2(\Omega_T)}$$
$$+ \left\langle -\nabla \cdot (\varsigma \nabla u), \lambda \right\rangle_{L^2(\Omega_T)}$$
$$+ \left\langle v, \mathcal{M}^* (\mathcal{M}u - d) \right\rangle_{L^2(\Omega_T)}$$
$$+ \left\langle \rho \frac{\partial^2 v}{\partial t^2} - \nabla \cdot (\mu \nabla v), \lambda \right\rangle_{L^2(\Omega_T)}$$

In order to cancel the last two terms we integrate by parts the last term to obtain the adjoint IBVP for λ

$$\rho(x)\frac{\partial^2 \lambda(x,t)}{\partial t^2} - \nabla \cdot (\mu(x)\nabla \lambda(x,t))$$
$$= -\mathcal{M}^* (\mathcal{M}u - d), \quad (x,t) \in \Omega \times (0,T).$$
$$\lambda(x,T) = \lambda_t(x,T) = 0, \qquad x \in \Omega.$$
$$\lambda(x,t) = 0, \qquad (x,t) \in \Gamma_1 \times [0,T].$$
$$\frac{\partial \lambda(x,t)}{\partial n} = 0, \qquad (x,t) \in \Gamma_2 \times [0,T].$$

Then

$$DJ(\rho, \mu)(\xi, \varsigma) = \left\langle \xi \frac{\partial^2 u}{\partial t^2}, \lambda \right\rangle_{L^2(\Omega_T)}$$
$$+ \left\langle -\nabla \cdot (\varsigma \nabla u), \lambda \right\rangle_{L^2(\Omega_T)}$$

Again, integrating by parts the last term

$$DJ(\rho, \mu)(\xi, \varsigma) = \left\langle \xi, \lambda \frac{\partial^2 u}{\partial t^2} \right\rangle_{L^2(\Omega_T)}$$

$$+ \left\langle \varsigma, \nabla \lambda \cdot \nabla u \right\rangle_{L^2(\Omega_T)}$$

$$= \left\langle (\xi, \varsigma), \int_0^T \left(\lambda \frac{\partial^2 u}{\partial t^2}, \nabla \lambda \cdot \nabla u \right) dt \right\rangle_{L^2(\Omega) \times L^2(\Omega)}.$$

Hence

$$\nabla J((\rho, \mu)) = \int_0^T \left(\lambda \frac{\partial^2 u}{\partial t^2}, \nabla \lambda \cdot \nabla u \right) dt$$

$$= \int_0^T \left(-\frac{\partial \lambda}{\partial t} \frac{\partial u}{\partial t}, \nabla \lambda \cdot \nabla u \right) dt.$$

7. Final Comments

It is apparent that the discrete framework introduced in this paper can be both, parallelized and extended to 2D and 3D. Regarding the regularization of the inverse problem, it is well known that the parameters have large variances, and the development of strategies to incorporate *a priori* information into the regularization constitute a trendy field of research. On the other hand, the discrete framework introduced in this paper allowed us to recover discontinuities on the density in the acoustic approximation. This result alone highlights the adequacy and our approach. We plan to report a Bayesian analysis of this problem elsewhere.

References

[1] A. Bamberger, G. Chavent, P. Lailly: *About the stability of the inverse problem in 1-d wave equations application to the interpretation of seismic profiles*. Applied Mathematics and Optimization **5** (1979), no. 1, pp. 1–47.

[2] J. Dieudonné: *Foundations of modern analysis*. Academic Press, New York and London (1960)

[3] C. Burstedde and O. Ghattas: *Algorithmic strategies for full waveform inversion: 1D experiments*. Geophysics **74**, no. 6 (2009), pp. 37–46 .

[4] A. Fichtner: *Full seismic waveform modelling and inversion*. Springer (2010).

[5] A. Fichtner, H. P. Bunge, H. Igel: *The adjoint method in seismology:: I. Theory*. Physics of the Earth and Planetary Interiors **157**, no. 1-2 (2006), pp. 86–104 .

[6] P. Hursky, M. B. Porter, B. D. Cornuelle, W. S. Hodgkiss, W. A. Kuperman: *Adjoint modeling for acoustic inversion*. J. Acoust. Soc. Am. **115**, no. 2 (2004), pp. 607–619.

[7] V. Isakov: *Inverse problems for partial differential equations*. Springer Verlag, 2006.

[8] O. A. Ladyzhenskaya: *The boundary value problems of mathematical physics*. Appl. Math. Sci. 49 (1985).

[9] J. R. McLaughlin, J. R. Yoon: *Unique identifiability of elastic parameters from time-dependent interior displacement measurement*. Inverse Problems **20** (2004), pp. 25–45 .

[10] J. Nocedal, S. J. Wright: *Numerical optimization*. Springer-Verlag, New York (1999).

[11] F. Ortigosa, M. A. Polo, F. Rubio, M. Hanzich, R. de la Cruz, J. M. Cela: *Evaluation of 3D RTM on HPC Platforms*. SEG Technical Program Expanded Abstracts, vol. 27 (2008), pp. 2879–2883 .

[12] R. E. Plessix: *A review of the adjoint-state method for computing the gradient of a functional with geophysical applications*. Geophys. J. Int. **167**, no. 2 (2006), pp. 495–503 .

[13] P. Sacks, F. Santosa: *A simple computational scheme for determining the sound speed of an acoustic medium from its surface impulse response*. SIAM J. Sci. Stat. Comput. 8, no. 4. (1987), pp. 501–520.

[14] W. W. Symes: *Stable solution of the inverse reflection problem for a smoothly stratified elastic medium*. SIAM J. Math. Anal. 12, no. 3 (1981), pp. 421–453 .

[15] W. W. Symes: *The seismic reflection inverse problem*. Inverse Problems **25** (2009).

[16] J. Virieux, S. Operto: *An overview of full-waveform inversion in exploration geophysics*. Geophysics **74** (2009).