



# A Two-Phase Approach for Full Waveform Inversion using Constrained Optimization

Alexandre William Camargo (IMECC/UNICAMP) & Lúcio Tunes Santos (IMECC/UNICAMP), University of Campinas, Brazil

Copyright 2019, SBGf - Sociedade Brasileira de Geofísica.

This paper was prepared for presentation at the 16<sup>th</sup> International Congress of the Brazilian Geophysical Society, held in Rio de Janeiro, Brazil, August 19-22, 2019.

Contents of this paper were reviewed by the Technical Committee of the 16<sup>th</sup> International Congress of The Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of The Brazilian Geophysical Society is prohibited.

## Abstract

**Most Full Waveform Inversion (FWI) formulations use Quasi-Newton optimization algorithms to perform the inversion for physical parameters. In this paper we are concerned with the inversion of the velocity model, where we propose a two-phase approach. In Phase I, we apply an Augmented Lagrangian method to a constrained optimization problem, where the constraints are based on velocity profiles and we consider that the velocity model can be parameterized by Radial Basis Functions. Moreover, the initial approximation for the velocity is a homogeneous model. In Phase II we apply the Spectral Projected Gradient method for a problem with box constraints only and without parameterization of the velocity model. Numerical experiments prove the efficiency of the new approach, and show the importance of Phase I as the output for a good (initial) approximation for the velocity model in Phase II, which is very similar to the classical formulation for FWI.**

## Introduction

The seismic Full Waveform Inversion (FWI) is a powerful tool seismic imaging since it tries to estimate subsurface parameters from seismic data. Theoretically, the FWI achieves a high resolution for the velocity model if the initial approximation is sufficiently close to the true (unknown) velocity model. The fundamental basis for FWI unites two important worlds of computational Geophysics that are seismic modeling and seismic inversion. The state of the art for classical FWI can be found in Lailly (1983), Tarantola (1984) and Virieux and Operto (2009). The classical 2D optimization problem in FWI is formulated as a curve fitting in the least squares sense, i.e.,

$$\text{Minimize } \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T dt |u_s(m; \mathbf{x}_r, t) - d_s(\mathbf{x}_r, t)|^2, \quad (1)$$

where  $m \equiv m(\mathbf{x}) = 1/c(\mathbf{x})^2$  is the square of slowness,  $c(\mathbf{x})$  is the acoustic velocity,  $\mathbf{x} = (x, z) \in \mathbb{R}^2$  is the spatial variable,  $u_s(m; \mathbf{x}_r, t) \equiv u(m; \mathbf{x}_s, \mathbf{x}_r, t)$  is the acoustic wavefield that satisfies, in this work, the bidimensional acoustic wave equation with constant density,  $d_s(\mathbf{x}_r, t) \equiv d(\mathbf{x}_s, \mathbf{x}_r, t)$  is the observed seismic data for time  $t \in [0, T]$ , with  $N_s$  source positions at  $\mathbf{x}_s = (x_s, z_s)$  and  $N_r$  receiver positions at  $\mathbf{x}_r = (x_r, z_r)$ .

However, optimization algorithms based on first order approximations require a good initial approximation for the velocity model, in order to guarantee the convergence to the true velocity model. This problem has been one of the big challenges for the success of the FWI and some alternatives to obtain a good initial approximations can be found applying tomography and stereotomography methods. See, e.g., Prioux et al. (2012) and Tavakoli et al. (2017).

The most established method that has been used for solving the classical formulation for the FWI is the quasi-Newton method L-BFGS (Limited memory Broyden-Fletcher-Goldfarb-Shanno) proposed by Liu and Nocedal (1989). However, the L-BFGS method solve an unconstrained nonlinear problem and, then, if there exist some additional information, it must be used as regularization and/or penalization terms. Some works using regularization and penalization formulations can be found in Guitton et al. (2012), Leeuwen and Herrmann (2016) and Esser et al. (2018).

## Constrained Optimization Problem for FWI

The constrained optimization problem that we wish to solve for the FWI is

$$\begin{aligned} &\text{Minimize } f(\mathbf{w}) \\ &\text{subject to } h(\mathbf{w}) = 0, \\ &g(\mathbf{w}) \leq 0, \\ &\mathbf{w} \in \Omega, \end{aligned} \quad (2)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h: \mathbb{R}^n \rightarrow \mathbb{R}^{N_h}$ ,  $g: \mathbb{R}^n \rightarrow \mathbb{R}^{N_g}$  are continuous functions and admit continuous first derivatives on  $\mathbb{R}^n$ , and  $\Omega \subset \mathbb{R}^n$  is closed.

The Augmented Lagrangian function is defined by

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = &f(\mathbf{w}) + \frac{\rho}{2} \|h(\mathbf{w}) + \boldsymbol{\lambda}/\rho\|^2 \\ &+ \frac{\rho}{2} \|(g(\mathbf{w}) + \boldsymbol{\mu}/\rho)_+\|^2, \end{aligned} \quad (3)$$

where  $\rho > 0$  is the penalty parameter,  $\boldsymbol{\lambda} \in \mathbb{R}^{N_h}$  and  $\boldsymbol{\mu} \in \mathbb{R}_+^{N_g}$  are approximations for the Lagrange multipliers associated equality and inequality constraints, respectively,  $[(y)_+]_i = \max\{0, y_i\}$ , and  $\|\cdot\|$  is the euclidian norm. Notice that  $\mathcal{L}(\mathbf{w}, \mathbf{0}, \mathbf{0})$  is the classical  $L_2$  penalty function. The Augmented Lagrangian algorithm for solving (2) proceeds by approximately minimizing (3) with respect to  $\mathbf{w}$  at each iteration, and updating the Lagrange multiplier vectors  $\boldsymbol{\lambda}$  and  $\boldsymbol{\mu}$ , and the penalty parameter  $\rho$ .

For the FWI problem, we consider that the acoustic propagation velocity can be written as a combination of some base functions,  $\phi_i: \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $i = 0, 1, \dots, n$ , i.e., the acoustic propagation velocity  $c$  is given by

$$c(\mathbf{w}; \mathbf{x}) = \phi_0(\mathbf{x}) + \sum_{k=1}^n w_k \phi_k(\mathbf{x}). \quad (4)$$

This is done to reduce the cost of the optimization procedure, since numerical tests considering all the points on the discretized model in the Augmented Lagrangian solver resulted in unviable CPU times, even for a small model for a serial implementation, i.e., without parallel computing. The base functions  $\phi_i(\mathbf{x})$  are chose according to desired complexity of the velocity model. For example, for a constant vertical gradient velocity model we can choose  $n = 2$  and  $\phi_0(\mathbf{x}) = c_0$ ,  $\phi_1(\mathbf{x}) = x$ , and  $\phi_2(\mathbf{x}) = z$ . Other simples choices are possible and can be found in Camargo and Santos (2017) and Camargo (2019).

Let us assume that we know the velocity somewhere in the analysed region (this information may come from a vertical well profile, for example):  $v(\mathbf{x}_q)$  for  $q = 1, \dots, N_q$ . Then, the constraints can be formulated as

$$h_q(\mathbf{w}; \mathbf{x}_q) = c(\mathbf{w}; \mathbf{x}_q) - v(\mathbf{x}_q), \quad (5)$$

or

$$g_q(\mathbf{w}; \mathbf{x}_q) = (c(\mathbf{w}; \mathbf{x}_q) - v(\mathbf{x}_q))^2 - \varepsilon v(\mathbf{x}_q)^2, \quad (6)$$

for  $q = 1, \dots, N_q$  and  $\varepsilon > 0$ .

### Radial Basis Function

For geological models with higher complexity, the velocity model can be parameterized using Radial Basis Functions (RBF) (Buhmann, 2003; Wendland, 2005). In equation (4), let us consider  $\phi_k(\mathbf{x}) \equiv \phi(\mathbf{x} - \mathbf{y}_k) = \varphi(\|\mathbf{x} - \mathbf{y}_k\|)$ , where  $\mathbf{y}_k \in \mathbb{R}^2$ ,  $k = 0, 1, \dots, n$ , are called centers, without any special assumptions of mesh regularity, and  $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}$ . Let us assume that at arbitrary points  $\mathbf{x}_\ell \in \mathbb{R}^2$ ,  $\ell = 0, 1, \dots, L$ , the corresponding values  $c(\mathbf{x}_\ell)$  are fixed, with  $L \geq n$ .

In the case of  $L = n$  and assuming that there is no noise in the data, we can interpolate  $c(\mathbf{x})$  via the linear combination given by equation (4). The coefficients  $\mathbf{w}$  are obtained by the solution of the linear system

$$A\mathbf{w} = \mathbf{b}, \quad (7)$$

where  $A$  is the square matrix  $L \times n$  with elements  $a_{i,j} = \phi_j(\mathbf{x}_i)$ ,  $i = 1, 2, \dots, L$ ,  $j = 1, 2, \dots, n$ , and  $\mathbf{b}$  is the vector with entries  $b_i = c(\mathbf{x}_i) - \phi_0(\mathbf{x}_i)$ ,  $i = 1, 2, \dots, L$ . For the RBF used in this work it can be proved that  $A$  is nonsingular.

In the case that the data set is contaminated with noise or  $L > n$ , instead of using interpolation, we can solve system (7) in the least square sense, i.e., the coefficients  $\mathbf{w} \in \mathbb{R}^n$  are given by  $\min_{\mathbf{w} \in \mathbb{R}^n} \|A\mathbf{w} - \mathbf{b}\|^2$ . This case will be denoted by LS-RBF.

### Two-Phase Approach for FWI

The Two-Phase approach for FWI can be summarized as:

**Phase I:** Solve the problem in standard form given by (2) using the Augmented Lagrangian algorithm proposed by Birgin and Martínez (2014). The acoustic wave equation is solved in time domain by a finite-difference method (LeVeque, 2007).

**PHASE II:** Solve a problem similar to the one given by equation (1), but in frequency domain and adding box constraints, and using the Spectral Gradient Projected (SPG) method proposed by Birgin and Martínez (2002), which is also is used for solving the subproblem associated to the minimization of (3). The acoustic wave equation is

solved in frequency domain (Helmoltz equation) (Leeuwen and Herrmann, 2016), for some values of the frequency.

Firstly, we choose the parameterization of the velocity model, define an initial approximation for  $\mathbf{w}$ , and perform **Phase I**. The final estimated model recovered by the parametrization in this phase is then used as an initial approximation for the velocity model in **Phase II**. Before of declaring the end of execution, we can decide to improve the information required in the parameterization (the RBF or LS-RBF nodes, for example). In the affirmative case we return to **Phase I**. Otherwise, we finish the execution.

### Types of Problems for Constrained FWI in Phase I

#### Problem 1

Using the RBF parameterization, the first constrained optimization problem is given by

$$\begin{aligned} &\text{Minimize} && f(\mathbf{w}) = J(\mathbf{w}) + \theta \|A\mathbf{w} - \mathbf{b}\|^2 \\ &\text{subject to} && h(\mathbf{w}) = 0, \\ &&& g(\mathbf{w}) \leq 0, \\ &&& \mathbf{w} \in \Omega, \end{aligned} \quad (8)$$

where  $\Omega = \{\mathbf{w} \in \mathbb{R}^n \mid -10^{21} \leq \mathbf{w}_i \leq 10^{21}, i = 1, \dots, n\}$ ,  $J: \mathbb{R}^n \rightarrow \mathbb{R}_+$  is a function that measures the misfit between observed and modeled data,  $\theta$  is a regularization parameter, and  $h$  and  $g$  are given by equations (5)–(6).

Based on Guitton (2012) the regularization parameter is given by

$$\theta = \frac{1}{2} \frac{J(\mathbf{w}^0)}{\|A\mathbf{w}^0 - \mathbf{b}\|^2}, \quad (9)$$

where  $\mathbf{w}^0$  is the initial guess for the inversion process.

Observe that the interpolation condition could be a linear equality constraint instead of being added to the misfit function. Numerical tests considering such possibility indicated that it may cause convergence to stationary points of the infeasibility measure,  $\|h(\mathbf{w})\|^2 + \|g(\mathbf{w})_+\|^2$ .

#### Problem 2

The second optimization problem uses the LS-RBF parameterization, which results in an unusual FWI formulation, since the functional  $J$  appears as an inequality constraint:

$$\begin{aligned} &\text{Minimize} && f(\mathbf{w}) = \|A\mathbf{w} - \mathbf{b}\|^2 \\ &\text{subject to} && h(\mathbf{w}) = 0, \\ &&& g(\mathbf{w}) \leq 0, \\ &&& J(\mathbf{w}) \leq \varepsilon J(\mathbf{w}^0), \\ &&& \mathbf{w} \in \Omega, \end{aligned} \quad (10)$$

where  $\Omega = \{\mathbf{w} \in \mathbb{R}^n \mid -100 \leq \mathbf{w}_i \leq 100, i = 1, \dots, n\}$ ,  $h$  and  $g$  are the equality and inequality constraints given by (5)–(6), and  $0 < \varepsilon < 1$ .

### Problem for Constrained FWI in Phase II

In Phase II, the constrained optimization problem is given by

$$\begin{aligned} &\text{Minimize} && \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{k=1}^{N_\omega} |\widehat{u}_s(\mathbf{m}; \mathbf{x}_r, \omega_k) - \widehat{d}_s(\mathbf{x}_r, \omega_k)|^2 + \gamma \|D(\mathbf{m})\|^2 \\ &\text{subject to} && \mathbf{m}_{min} \leq \mathbf{m}_i \leq \mathbf{m}_{max}, \quad i = 1, 2, \dots, N, \end{aligned} \quad (11)$$

where  $\gamma$  is a regularization parameter,  $N = N_x \times N_z$  is the total number of points for the spatial discretization,  $N_\omega$  is the number of chosen frequencies where  $\hat{u}_s$  is to be computed, and  $D$  is the forward first order difference operator.

### Misfit function and Gradient in Phase I

To analyse the sensitivity of the parameterization in the objective function, we choose to perform in Phase I tests using the following three choices for the misfit function  $J$ . Let us write

$$J(\mathbf{w}) = \int_{\mathbb{D}} d\mathbf{x} \int_0^T dt \mathcal{H}(u(\mathbf{w}; \mathbf{x}, t)), \quad (12)$$

where  $\mathbb{D} \subset \mathbb{R}^2$  is the spatial domain,  $T$  is the maximum recording time, and  $\mathcal{H}$  is one of the following functions:

#### 1. Weighted Least Squares (WLS),

$$\mathcal{H}(u) = \frac{1}{2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_r} \delta(\mathbf{x} - \mathbf{x}_{r_j}) \left[ \frac{u_i(\mathbf{w}; \mathbf{x}, t) - d_i(\mathbf{x}, t)}{\alpha_i} \right]^2; \quad (13)$$

#### 2. Weighted Least Squares with Offset Weighting (WLS-OW),

$$\mathcal{H}(u) = \frac{1}{2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_r} \delta(\mathbf{x} - \mathbf{x}_{r_j}) \left\{ \|\mathbf{x}_{s_i} - \mathbf{x}\| \left[ \frac{u_i(\mathbf{w}; \mathbf{x}, t) - d_i(\mathbf{x}, t)}{\alpha_i} \right] \right\}^2; \quad (14)$$

#### 3. Cauchy,

$$\mathcal{H}(u) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_r} \delta(\mathbf{x} - \mathbf{x}_{r_j}) \ln \left[ 1 + \left( \frac{u_i(\mathbf{w}; \mathbf{x}, t) - d_i(\mathbf{x}, t)}{\beta_{i,j}} \right)^2 \right]. \quad (15)$$

In the equations above,  $\delta$  is the Dirac's delta "function",

$$\alpha_i = \sqrt{\sum_{j=1}^{N_r} \langle d_i(\mathbf{x}_{r_j}, t), d_i(\mathbf{x}_{r_j}, t) \rangle}, \text{ for } i = 1, 2, \dots, N_s, \text{ and } \beta_{i,j} =$$

$$\sqrt{\frac{1}{T} \langle d_i(\mathbf{x}_{r_j}, t), d_i(\mathbf{x}_{r_j}, t) \rangle}, \text{ for } i = 1, 2, \dots, N_s \text{ and } j = 1, 2, \dots, N_r,$$

and  $\langle \cdot, \cdot \rangle$  denotes the canonic inner product,

$$\langle y_1(t), y_2(t) \rangle = \int_0^T dt y_1(t) y_2(t). \quad (16)$$

The gradient of  $J$  is computed by the adjoint-state method (Fichtner, 2011) and is given by

$$[\nabla J(\mathbf{w})]_i = -2 \int_{\mathbb{D}} \frac{\phi_i(\mathbf{x})}{c(\mathbf{w}, \mathbf{x})^3} I(\mathbf{x}) d\mathbf{x}, \quad i = 1, 2, \dots, n, \quad (17)$$

with

$$I(\mathbf{x}) = \int_0^T dt q(\mathbf{x}, T-t) u_{tt}(\mathbf{w}; \mathbf{x}, t), \quad (18)$$

where  $q$  is the adjoint-state of  $u$ .

## Numerical Experiments

### Phase I parameters

The seismic data was obtained by a finite-difference method (LeVeque, 2007) with a centered scheme of second order in time and sixth order in space. During

the inversion process the same scheme is used, but only fourth order in space. The point source is a Ricker wavelet with peak frequency of 20 Hz. The absorbing boundary conditions (Kosloff and Kosloff, 1986) has an effective length of 51 points on all edges of the model. We have used the solver ALGENCAN (Birgin and Martínez, 2014) with tolerance  $\varepsilon_{opt} = 10^{-4}$  for stationary points, feasibility tolerance  $\varepsilon_{feas} = 10^{-5}$ , maximum number of inner iterations equals to 20 in both problems and maximum number of external iterations equals to 16 in Problem 1 and 8 in Problem 2. In both problems, the initial approximation for the Lagrange multipliers is the null vector. For the RBF, we used the Wendland-2 function and the centers were chosen through the Halton's sequence (Halton, 1960) using the velocity model illustrated in Figure 1, where the black circles indicate a number of 251 centers used in Problem 1, i.e.,  $\mathbf{w} \in \mathbb{R}^{251}$ .

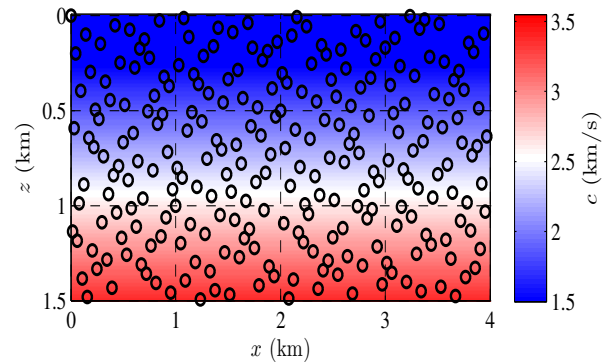


Figure 1: Velocity model used to choose the centers of the RBF by the Halton's sequence. The knots containing the velocity information are obtained by the nearest neighbors of centers.

For Problem 2, the number of centers was reduced to 51, that is,  $\mathbf{w} \in \mathbb{R}^{51}$ . The collected knots containing the velocity data set are obtained by the nearest neighbors of centers. The background function in the linear combination was  $\phi_0(\mathbf{x}) = 1.5$  (in km/s). The target model is the Marmousi II (Martin et al., 2002) shown in Figure 2, which contains three velocity profiles used in the constraints, with random noise of 10% on the bigger velocity, where each one has 31 points. In the experiments, we consider the first 23 points for the equality constraints and the remaining 70 for the inequality constraints, with  $\varepsilon = 0.001 v_{max}$ , where  $v_{max} = \max_{1 \leq q \leq 93} v(\mathbf{x}_q) = 3.7$  km/s.

The acquisition survey is composed by 5 common-shots with 101 receivers each, where the first shot is located at  $\mathbf{x}_s = (0.5 \text{ km}, 0.25 \text{ km})$ , with source spacing of 500 m, minimum offset of 200 m, and receiver spacing of 10 m.

### Phase II parameters

In Phase II, the seismic acquisition line is the same as in Phase I, but with a denser distribution of shots and receivers: 86 shots with spacing of 40 m, the first shot located at  $(0.1 \text{ km}, 0.25 \text{ km})$  and 172 receivers distributed along the range 100 m to 3.52 km with spacing of 20 m. We perform the seismic inversion in frequency domain, i.e., we solve Helmholtz's equation for modeling at frequencies 4 Hz, 5 Hz, 7 Hz, 12 Hz and 16 Hz. These values were based on the work of Tavakoli et al. (2017). The optimization

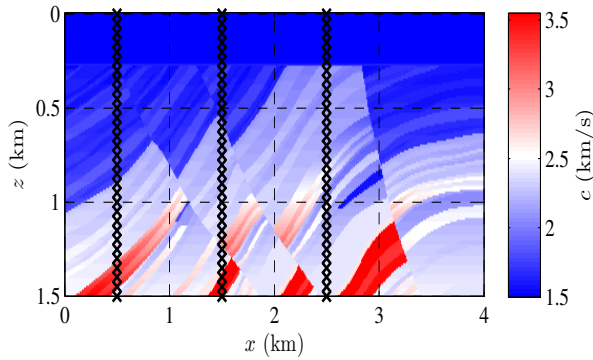


Figure 2: Marmousi II model. The  $\times$  indicates the positions for the velocity profiles.

was done using the SPG method (Birgin and Martínez, 2002) with optimization tolerance  $\varepsilon = 10^{-8}$ , and maximum number of iterations equals to 40. Moreover,  $N = 60,551$ ,  $\mathbf{m}_{min} = 1/v_{max}^2 = 1/3.7^2 = 0.07$  [s/km] $^2$ , and  $\mathbf{m}_{max} = 1/1.5^2 = 0.44$  [s/km] $^2$ .

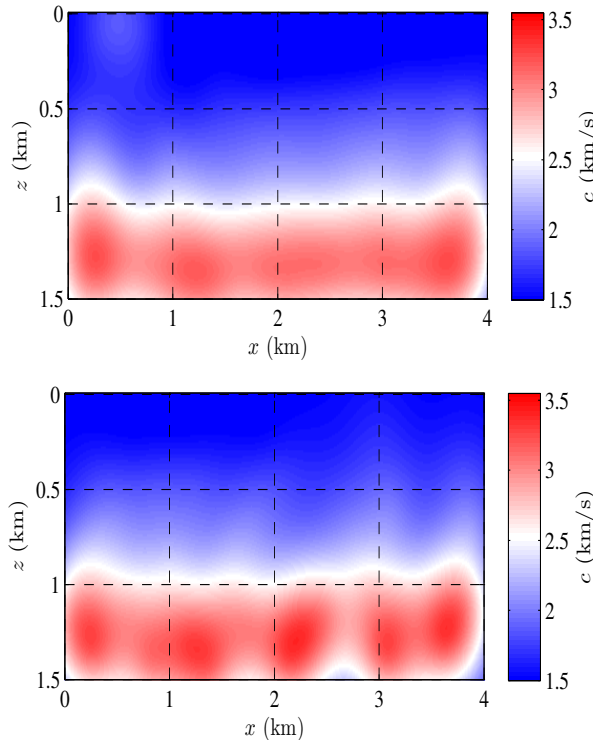


Figure 3: Final approximation in Phase I resulting from Problem 1 (top) and Problem 2 (bottom).

### Numerical analysis

The initial approximation in Phase I in both problems is  $\mathbf{w}^0 = 0$ , i.e., a homogeneous velocity model  $c(\mathbf{w}^0, \mathbf{x}) = 1.5$  km/s. Figure 3 shows the approximations obtained by ALGENCAN using Cauchy's functional (15), for Problems 1 and 2, respectively, since of the three functional analysed in the experiments, it is the one that presented more robust results. The algorithm stopped because it reached the maximum number of iterations. In Figure 4 are depicted the velocity profile at  $x = 1.5$  km compared with the profiles

obtained for each functional tested in Phase I, for Problems 1 and 2. Although the final approximations do not present a good resolution when compared with the target model, the misfit in the velocity profiles for both problems are in good agreement, which contributes to a good initial approximation for Phase II.

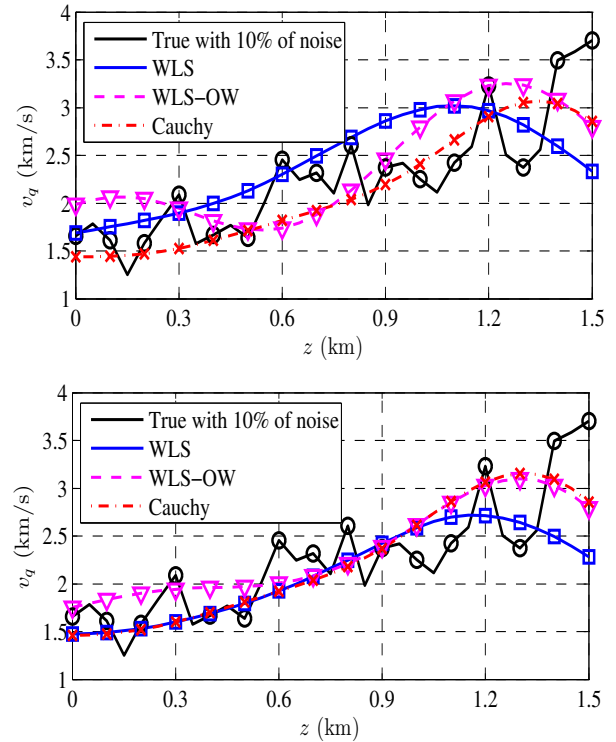


Figure 4: Comparison of velocity profiles at  $x = 1.5$  km resulting from Problem 1 (top) and Problem 2 (bottom).

Figure 5 shows the final approximations obtained in Phase II, using the approximation obtained in Phase I and Figure 6 shows the estimated model for an initial affine velocity model, that is, the same velocity model used for the choice the RBF centers (Figure 1). These results will be used as a reference model for comparison with the Two-Phase approach. Comparing with Figure 5, we observe that the Two-Phase approach has a better resolution in the region  $[1, 3] \times [1, 1.5]$ . Observing the bottom of Figure 5, we conclude that in the region  $[3, 4] \times [1, 1.5]$ , Problem 2 had a worse resolution.

In order to analyze the robustness of the Two-Phase approach, Figure 7 shows the relative errors for the estimated velocity in Phase II, using Problem 1, Problem 2 (using the initial model obtained in Phase I), and the affine velocity model (see Figure 1) as the initial approximation (without Phase I).

For the case of Problem 1, less than 13% of the grid points present a relative error greater than 15%, whereas for the other two cases, this percentage increases to 20%. However, these values can be reduced once we update the data used in the RBF, and perform Phase I & II again, since these final approximations are of better quality than the model used previously (Figure 1).



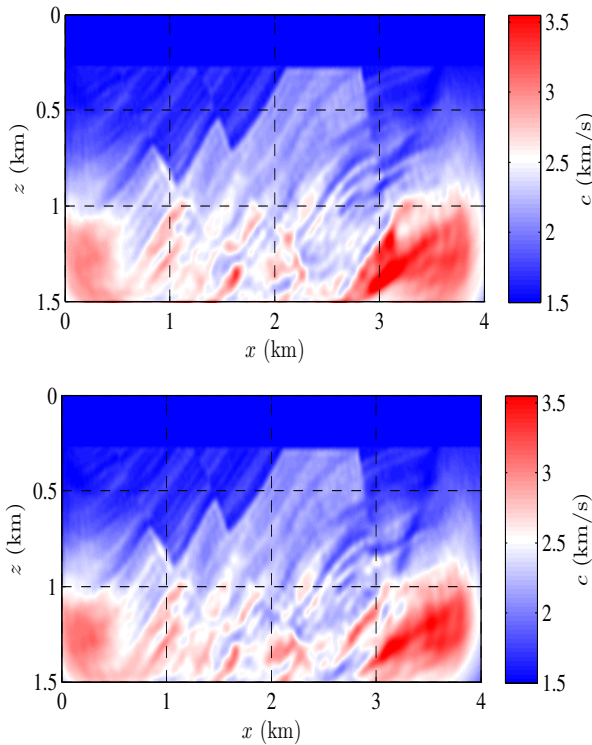


Figure 5: Final approximation in Phase II with initial approximation from Problem 1 (top) and Problem 2 (bottom).

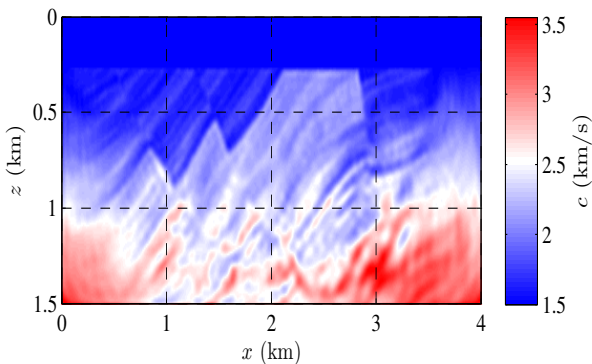


Figure 6: Final approximation in Phase II with initial approximation from affine velocity model without using Phase I.

## Conclusions

In this work, we present a Two-Phase approach for FWI, based on constrained optimization problems. In Phase I, the constraints are constructed using velocity profiles and, in Phase II, they are simple box constraints. In addition, in Phase I it is used a parameterization of the velocity model based on Radial Basis Functions, thus reducing the number of inversion variables. The numerical experiments show that we can start in Phase I with a constant velocity model thanks, to the profile information, and the final velocity model obtained becomes a good initial guess for the point-to-point inversion in Phase II. Besides that, this approach is better alternative than the application of tomography and/or stereotomography methods to find

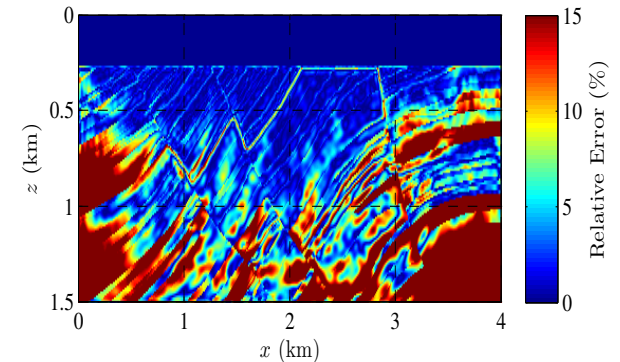
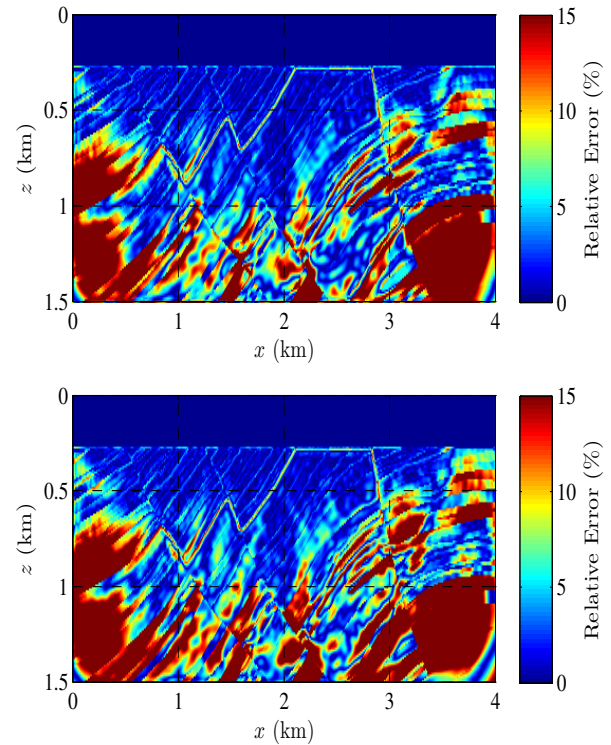


Figure 7: Relative error for Problem 1 (top), Problem 2 (middle), and reference model [affine velocity] (bottom).

an initial velocity model for FWI, since the Two-Phase approach is robust enough and the use of RBF provides the possibility of smart upgrades in the initial data.

## Acknowledgments

This work was kindly supported by the Brazilian research agencies CAPES - financing code 001 and CNPq (140893/2016-4; 311358/2017-9), and the sponsors of the Wave Inversion Technology (WIT) Consortium, Germany.

## References

- Birgin, E. G., and J. M. Martínez, 2002, A large-scale active-set box-constrained optimization method with spectral projected gradients: *Computational Optimization and Applications*, **23**, 101–125.
- , 2014, *Practical augmented lagrangian methods for constrained optimization*: Society for Industrial and Applied Mathematics.
- Buhmann, M. D., 2003, *Radial basis functions: theory*

- and implementation: Cambridge University Press. Cambridge Monographs on Applied and Computational Mathematics.
- Camargo, A. W., 2019, Análise da inversão da forma de onda completa pelo método do lagrangiano aumentado: PhD thesis, University of Campinas.
- Camargo, A. W., and L. T. Santos, 2017, Application of an augmented lagrangian method for constrained full waveform inversion: *WIT Annual Report*, **21**, 21–34.
- Esser, E., L. Guash, T. V. Leeuwen, A. Y. Aravkin, and F. J. Herrmann, 2018, Total-variation regularization strategies in full-waveform inversion: *SIAM Journal on Imaging Sciences*, **11**, 376–406.
- Fichtner, A., 2011, Full seismic waveform modeling and inversion: Springer. *Advances in Geophysical and Environmental Mechanics and Mathematics*.
- Guitton, A., 2012, Blocky regularization schemes for full waveform inversion: *Geophysical Prospecting*, **60**, 870–884.
- Guitton, A., G. Ayeni, and E. Díaz, 2012, Constrained full-waveform inversion by model reparametrization: *Geophysics*, **77**, R117–R127.
- Halton, J. H., 1960, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integral: *Numerische Mathematik*, **2**, 84–90.
- Kosloff, R., and D. Kosloff, 1986, Absorbing boundaries for wave propagation problems: *Journal of Computational Physics*, **63**, 363–373.
- Lailly, P., 1983, The seismic inversion problem as a sequence of before stack migrations: Conference on Inverse Scattering: Theory and Application, Society for Industrial and Applied Mathematics, 206–220.
- Leeuwen, T. V., and F. J. Herrmann, 2016, A penalty method for pde-constrained optimization in inverse problems: *Inverse Problems*, **32**, 1–27.
- LeVeque, R. J., 2007, Finite difference methods for ordinary and partial differential equations: Steady-state and time dependent problems: Society for Industrial and Applied Mathematics.
- Liu, D. C., and J. Nocedal, 1989, On the limited memory bfgs method for large scale optimization: *Mathematical Programming*, **45**, 503–528.
- Martin, G. S., K. J. Marfurt, and S. Larsen, 2002, Marmousi 2: An updated model for the investigation of avo in structurally complex areas, in *SEG Technical Program Expanded Abstracts 2002*: Society of Exploration Geophysicists, 1979–1982.
- Prieux, V., G. Lambaré, S. Operto, and J. Virieux, 2012, Building starting models for full waveform inversion from wide-aperture data by stereotomography: *Geophysical Prospecting*, **209**, 1629–1647.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- Tavakoli, B., S. Operto, A. Ribodetti, and J. Virieux, 2017, Slope tomography based on eikonal solvers and the adjoint-state method: *Geophysical Journal International*, **209**, 1629–1647.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, 127–152.
- Wendland, H., 2005, Scattered data approximation: Cambridge University Press. Cambridge Monographs on Applied and Computational Mathematics.