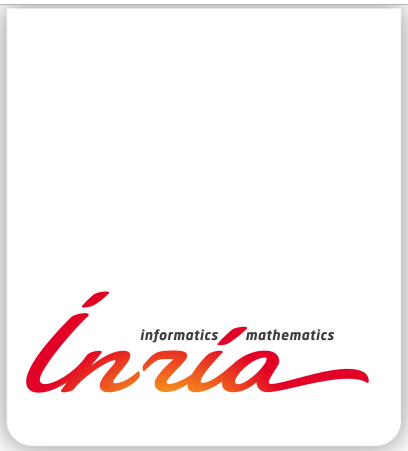


A priori estimates of attraction basins for velocity model reconstruction by time-harmonic Full Waveform Inversion and Data Space Reflectivity formulation

Hélène Barucq, Henri Calandra, Guy Chavent, Florian Faucher

RESEARCH REPORT
N° 9253
February 2019
Project-Team Magique-3D

ISSN 0249-6399 ISRN INRIA/RR--9253--FR+ENG



A priori estimates of attraction basins for velocity model reconstruction by time-harmonic Full Waveform Inversion and Data Space Reflectivity formulation

Hélène Barucq*, Henri Calandra†, Guy Chavent‡, Florian Faucher*

Project-Team Magique-3D

Research Report n° 9253 — February 2019 — 51 pages

Abstract: Determination of background velocity by Full Waveform Inversion is known to be hampered by the local minima of the data misfit caused by the phase shifts associated to background perturbations. We give in this paper an a priori study of the *convergence basins* around any nominal velocity model, where the data misfit is guaranteed to have only one local - and hence global - minimum, provided the data error is below a given *tolerable error level*. The size of the attraction basin gives an upper bound to the error in the initial velocity model which ensures that local optimization algorithm will converge to the global mimimum. We use geometry in the data space in order to compute this size and the associated tolerable error *without having to solve any FWI least squares minimization problem*. Application is made to the inverse problem associated with wave equations in the time-frequency domain, using reflection data. We study first how, when FWI is used, frequency and search direction geometry influence the size of attraction basins, and how to foster it (e.g. complex frequency progression). Then, we compare the size of attraction basins for background velocities for the classical FWI and the Data Space Reflectivity MBTT formulation. We investigate the domain of validity of the MBTT reformulation, and show that it can increase the size of the background attraction basins by a factor of four to fifteen.

Key-words: Inverse problem, optimization, convergence analysis, geophysics, wave equations, FWI, MBTT, Helmholtz equation, reconstruction algorithm.

* Project-Team Magique-3D, Inria Bordeaux Sud-Ouest, Université de Pau et des pays de l'Adour, France. (florian.faucher@inria.fr).

† Total Research Exploration & Production, Houston, TX 77002, USA.

‡ Project-Team Serena, Inria Paris Research Center, France.

RESEARCH CENTRE
BORDEAUX – SUD-OUEST

200 avenue de la Vieille Tour
33405 Talence Cedex

Estimations a priori de bassins d'attraction pour la reconstruction de modèles de vitesse pour l'équation de Hemholtz par inversion de forme d'onde (FWI) et par la formulation stockant la réflectivité dans l'espace des données (MBTT)

Résumé : La détermination des vitesses de background par inversion de forme d'onde (FWI - *Full Waveform Inversion*) est perturbée par les minima locaux présents dans la fonction coût, qui sont dus aux déphasages engendrés par les perturbations de ce background. Dans ce papier, nous donnons une étude à priori des *bassins de convergence* autour d'un modèle de vitesse nominal, pour lequel la fonction coût est garantie de n'avoir qu'un seul local - et donc global - minimum, tant que l'erreur sur les données est inférieure a un certain *niveau d'erreur tolérable*. La taille du bassin d'attraction donne une borne supérieure de l'erreur permise sur le modèle initial pour garantir qu'un algorithme d'optimisation local converge vers un minimum global. Nous utilisons des outils géométriques dans l'espace des données pour calculer la taille du bassin, et l'erreur tolérable associée, *sans avoir à résoudre un quelconque problème de minimisation.* Nous appliquons la méthodologie au problème inverse pour l'équation des ondes, dans le domaine fréquentiel, avec des données de réflexion. Nous étudions d'abord comment, avec la FWI, la taille du bassin d'attraction est influencée par la fréquence et la direction de recherche, ainsi que les améliorations que l'on peut y apporter (avec des fréquences complexes par exemple). Ensuite, nous comparons la taille du bassin d'attraction pour la reconstruction des vitesses de background entre l'approche FWI classique et sa reformulation stockant la réflectivité dans l'espace des données (méthode MBTT - *Migration Based Travel Time*). Nous étudions enfin le domaine de validité de la formulation MBTT, et montrons qu'elle peut mener à un gain d'un facteur quatre à quinze sur la taille du bassin.

Mots-clés : Problème inverse, optimisation, analyse de convergence, géophysique, équation des ondes, FWI, MBTT, algorithme de reconstruction.

Contents

1 Introduction **4**

2 Time-harmonic inverse wave problem **5**
2.1 Time-harmonic wave equations 5
2.2 Model and data spaces 6
2.3 Inversion via classical FWI 7
2.4 Inversion via MBTT–FWI (*background/data-space-reflectiv*ity decomposition) 8

3 Optimizability and Attraction Basins **9**
3.1 Global Radius of Curvature and Deflection 10
3.2 Directional Attraction Basins 11
3.3 Local Θ-estimate for $\Delta^{\mathfrak{u}}_{m_0}$ and associated tolerable error level $R^{\mathfrak{u}}_{m_0}$ 12
3.4 Exact Θ- and R_G-estimates of $\Delta^{\mathfrak{u}}_{m_0}$ and associated tolerable error levels 13

4 FWI: Numerical determination of convergence basins and tolerable error levels **14**
4.1 Influence of frequency ω and search direction $\mathfrak{u}$ 14
4.2 Influence of the initial model 17
4.3 Approximate n-dimensional attraction basins 18
4.4 Frequency bandwidth data 20
4.5 Selection of complex frequencies 21

5 Optimizability: MBTT versus classical FWI **24**
5.1 Choice of the model 24
5.2 Choice of perturbation directions 26
5.3 Comparison of Local Θ-estimates 27
5.4 Comparison of exact Θ- and R_G-estimates 27
5.5 Influence of reflectivity level 30
5.6 Influence of background smoothness 31
5.7 Behavior of the data misfit 32
5.8 Strategy for MBTT waveform inversion 34

6 Extension for other applications of least squares minimization **35**
6.1 Elastic medium reconstruction 35
6.2 Extension to different boundary conditions 36

7 Conclusion **39**

A Adjoint state using complex variables **40**
A.1 Complex derivation 40
A.2 Adjoint state method 42
A.3 Directional derivative computation 43

B MBTT (Data Space) model parameterization **44**
B.1 Expression of the reflectivity with MBTT 44
B.2 Directional derivative computation 45

C Influence of the geometry and frequency on FWI **45**
C.1 Comparison between salt and Marmousi geometries 45
C.2 Low frequencies compensation 46

1 Introduction

The Full Waveform Inversion (FWI) approach to seismic inversion consists in minimizing the misfit between the recorded and synthetic data with respect to the Earth parameters. It uses the full waveform of the recorded waves as data, which makes it free of horizon picking, and hence a good candidate for automatic inversion. It was first introduced by Bamberger et al. [5, 6] for one-dimensional model, and followed, in two dimensions, by Lailly [37] and Tarantola [61]. In particular, [37] made the link between the gradient of the misfit functional, computed by the adjoint state method, and the well-known migration imaging operator used by geophysicists. Early numerical experimentations in [62, 32] already show the strength and the weakness of FWI: it can produce very clear and detailed images of the sub-subsurface (strength), under the imperative condition that a good background velocity model is available (weakness). In fact, the misfit functional with respect to background parameters shows local minima due to phase shift/cycle skipping, which hamper the reconstruction by local algorithms, unless the minimization starts inside the attraction basin. FWI is extended to elastic waves in [43], with implementation in [26], to resolve the short wavelengths of P- and S-impedance. Initially carried out in the time-domain, FWI can also be used in the frequency domain, as promoted by Pratt and collaborators in [51, 48, 50]. The use of complex frequencies was later introduced by [53, 54], giving rise to the so-called Laplace and Laplace-Fourier domain FWI.

In the recent years, with the increase of the computational power, the solution of the FWI approach to seismic imaging by local optimization algorithms has been extensively developed. However, this approach does not solve the inherent problem of local minima: the determination of the full velocity model, including the low spatial frequency background, requires data with very low frequency (as unrealistic as 1 Hz), coupled with a resolution from low to high frequency to avoid 'cycle-skipping', as for example in [16, 59] (similar behaviours are studied in electromagnetism in [7, 8]). One could also resort to random optimization techniques e.g., [33], which have the ability to find the global minimum even in presence of many parasitic local minima. But they require a very large number of misfit evaluations for the determination of a relatively small number of parameters: they are not adapted for large number of unknowns (in our experiments, we consider several hundred thousands of parameters).

Different misfit functionals have been considered to alleviate the phase shift/cycle skipping problem, such as the logarithmic norm, explored by [62] and [57, 58]. [13] uses a criterion based on the envelope of the signal; [41, 52, 70] are based upon an optimal transport distance. Then, Newton type algorithms provide a natural framework to solve the nonlinear minimization problem in FWI, cf. [49]. However, they require the computation and inversion of the Hessian, which may be overwhelming due to the numerical cost. Thus, some algorithms simply account for the diagonal of the Gauss-Newton Hessian, e.g., [55]; multi-scale Newton methods are studied in [1] for the same reason. Conjugate gradient method with Hessian-vector computation is implemented in [42]. The estimation of the step length is another challenge for the iterative minimization problem and is often conducted using the line search method (e.g., [44]). Alternatives for a more precise step exist, such as the trust region method [29], the Maximum Projected Curvature method [19]. For Landweber iteration scheme, this step is further given in [27]. The choice of search direction and line search method naturally influences the performance of FWI algorithm, comparisons are given in [9] in the context of inverse scattering. Yet, none of these local optimization methods can avoid local minima.

In order to overcome the presence of local minima when solving the FWI problem by local optimization algorithms, alternative parametrizations have been proposed early. They are based on a decomposition of the Earth parameter into separate background and reflectivity parameters, and result in an increase of the computational complexity. The Differential Semblance Optimization (DSO) of Symes [60] extends the depth reflectivity model to account for the various illuminations in the data, defining a semblance criterion whose minimization produces the desired background model [60]; The Migration Based Travel Time (MBTT) approach [21, 24, 22] parameterizes the depth reflectivity by a *data-space reflectivity*, which is used to store the reflectivity during the update of the background. Note that the so-called Reflection FWI [69] retains from MBTT the migration-demigration process, but differs in that it does not use the data-space reflectivity concept, and uses as search direction the gradient of the classical FWI data misfit.

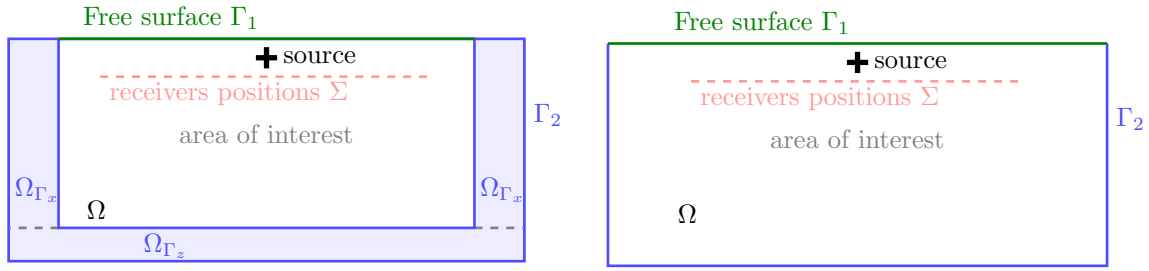
In this paper, we investigate the *optimizability of the FWI problem* associated with the time-

frequency domain wave equation, and compare it with that of the *MBTT reformulation* (in terms of background and data-space reflectivity). The objective is to *quantify the size of "attraction basins"* and *"tolerable level errors"* which ensure that a local algorithm with an initial guess inside the basin will converge to the global minimum, provided the error on the data is less than the tolerable error. By construction, these quantities depend solely on the (parameter-to-synthetic) forward map, but not on the data, and can be computed without any minimization of the data misfit.

Section 2 presents the time-harmonic wave equation to be inverted, together with the FWI minimization problem and its MBTT reformulation employed for the model parameters reconstruction. Section 3 presents the required geometrical tools of [19] for *optimizability* study. Sufficient and necessary optimizability conditions are given, and estimates of the size of the *attraction basin* in the parameter space, and of the *tolerable error level* in the data space are derived. Section 4 analyses the influence of (complex) frequency and search geometry on the size of attraction basins and tolerable error levels for the classical FWI formulation. In Section 5, we compare the optimizability of classical FWI and its MBTT reformulation. We expect that the migration-demigration included in MBTT eliminates phase shift/cycle skipping problems, leading to larger attraction basins for the minimum with respect to background, and investigate the effectiveness of this approach. In Section 6 we analyze the optimizability of least squares minimization algorithm for other context (elasticity, boundary conditions). In Appendix A, we review the gradient computation in the frequency domain, emphasizing the specificity of complex valued fields for the adjoint state method. The MBTT decomposition is specified in Appendix B. Appendix C provides numerical experiments of acoustic FWI to validate our estimates.

2 Time-harmonic inverse wave problem

We formulate the seismic inverse problem associated with the wave equation, for the recovery of acoustic or elastic materials from reflection data. We consider a bounded domain Ω of $\mathbb{R}^2$ which represents the region of interest in the subsurface (note that we restrict ourselves to two dimensions for the numerical experiments but the analysis holds similarly in three dimensions). The boundaries are denoted $\partial\Omega = \Gamma_1 \cup \Gamma_2$, where we distinguish the upper free surface (Γ_1) from the numerical boundary (Γ_2). We assume that partial data are acquired on a portion Σ of the domain, which possibly coincides with a part of the boundary. Figure 1 illustrates the configuration.



(a) Computational domain using Perfectly Matched Layers (PML) at the lateral and bottom boundaries.

(b) Computational domain using absorbing boundary condition on Γ_2.

Figure 1: Illustration of the two-dimensional area of interest. It characterizes a geophysical situation where sources and receivers are located in the upper part. The boundary is decomposed into a free surface at the top (Γ_1); on the sides and bottom, conditions (Γ_2) are implemented to avoid returning waves.

2.1 Time-harmonic wave equations

The propagation of waves is defined by the displacement field $\boldsymbol{u}$, solution to the following partial differential equation,

$$-\omega^2\rho(\mathbf{x})\boldsymbol{u}(\mathbf{x}) - \nabla\cdot\underline{\sigma}(\mathbf{x}) = \boldsymbol{g}(\mathbf{x}), \qquad \text{in } \Omega. \tag{1}$$

We have introduced the (possibly complex) angular frequency ω, the medium density ρ and the (Cauchy) stress tensor of order two $\underline{\sigma}$. The source is represented by $\boldsymbol{g}$. We disregard boundary conditions for now. The stress tensor encompasses the different material properties and reduces, for example, when considering isotropy (see [64, 65]).

In the case of acoustic propagation (i.e. propagation in fluid), the wave equation is established for the scalar pressure field p instead. It is based upon the Euler equations (see, for example, [25, 35]). If we further assume a homogeneous density, then the wave equation coincides with the Helmholtz equation (see, e.g., [31]), and the pressure field p is solution to

$$-\frac{\omega^2}{c^2(\mathbf{x})}p(\mathbf{x}) - \Delta p(\mathbf{x}) = h(\mathbf{x}), \qquad \text{in } \Omega, \tag{2}$$

where $c(\mathbf{x})$ is the velocity (wave speed) and h the scalar source. The upper part of the boundary, Γ_1 (see Figure 1), corresponds with a physical free surface, i.e. the interface between the air and the medium, the field satisfies

$$p(\mathbf{x}) = 0, \qquad \text{on } \Gamma_1 \text{ (free surface condition)}. \tag{3}$$

Due to the numerical truncation of the real domain (the Earth), appropriate conditions are imposed on Γ_2 (see Figure 1(b)). One possibility is the consideration of absorbing boundary condition (see [30]) to ensure that waves that reach Γ_2 are not reflected back to the domain, it is given by

$$\partial_\nu p(\mathbf{x}) - \mathrm{i}\omega c^{-1}(\mathbf{x})p(\mathbf{x}) = 0, \qquad \text{on } \Gamma_2 \text{ (absorbing boundary condition)}, \tag{4}$$

where ν indicates the normal direction. As an alternative, one can consider instead Perfectly Matched Layers (PML, see [10]) on the sides and bottom boundary (see Figure 1(a)) where the derivatives become

$$\partial_x \to \big(1 + \mathrm{i}\frac{\sigma(x)}{\omega}\big)^{-1}\partial_x, \qquad \text{in } \Omega_{\Gamma_x} \text{ (Perfectly Matched Layer)}, \tag{5}$$

and analogously for the other direction, on Ω_{Γ_z}. In our implementation, the damping function σ is defined following the work of [66, 68].

The inverse problem aims the recovery of the physical parameter $c(\mathbf{x})$ in (2), using observations of the wave phenomenon $p(\mathbf{x})$ at some location. In the following, we mainly work with this acoustic inverse problem but we also offer some analysis for the isotropic elastic situation in Subsection 6.1 and for other boundary conditions in Subsection 6.2. Note that the methodology developed below is not restricted to inverse wave problem or geophysical setup, and can be applied in any context involving least-squares minimization schemes.

2.2 Model and data spaces

We denote by $m \in \mathcal{M}$ the subsurface model parameter to be recovered, here $m = c^{-2}$ in (2), and by $\mathcal{M} \subset E$ the admissible parameter set, which is a subset of the parameter space E. Here, $E = \mathbb{R}^n$ where n indicates the number of unknowns in the representation of m. We assume that $\mathcal{M}$ is a closed convex subset of E equipped with the L^2 norm $\| \, . \, \|_E$:

$$\|m\|_E = \Big(\sum\nolimits_{i=1}^{n} m_i^2 \Big)^{1/2}, \tag{6}$$

where m_i is the i^{th} component of m.

Remark 1 (Model representation)**.** *For the computational resolution of the inverse problem in the discrete settings, it is natural to consider the parameter to be recovered with a piecewise constant representation, such that*

$$m(\boldsymbol{x}) = \sum_{i=1}^{n} m_i \chi(D_i). \tag{7}$$

Here, m_i, for $i = 1$ to n, are scalar coefficients (which represent the model) and $\chi(D_i)$ represents the characteristic function on the domain region D_i. The subdomains D_i (for $i = 1$ to n) are non-overlapping and verify

$$\Omega = \cup_{i=1}^{n} D_i, \qquad D_i \cap D_j = \emptyset, \quad \forall i \neq j. \tag{8}$$

Therefore the model is represented via n coefficients, and n is the number of unknowns in the FWI inverse problem. Practically, the representation usually corresponds with the space discretization employed for the resolution of the wave equation: one coefficient per cell for discretization based upon mesh (e.g., Finite Element, Discontinuous Galerkin) or one value per node for Finite Differences discretization. Moreover, the piecewise constant model representation (7) *is also fundamental to show properties of stability, see [3, 11] and Remark 4.*

The data are recorded by the receivers on a portion Σ of the domain Ω, they contain both direct and reflected energy. Because one aims to image the deep structures of the earth, the direct arrivals need to be eliminated from the recorded data before they can be used in the inversion procedure. This is automatically done when a 'linearized' model is used, but in general, for the full Helmholtz problem (2), it requires subtracting the direct arrivals. So we shall denote by $\boldsymbol{d}$ the data deprived from their first arrivals, and by $\mathcal{D}$ the associated data space. In the time-harmonic case, $\boldsymbol{d}$ is a vector of q complex numbers and $\mathcal{D} = \mathbb{C}^q$, with $q = n_{frequency} \times n_{source} \times n_{receiver}$. The data space $\mathcal{D}$ is equipped with the norm $\| \, . \, \|_{\mathcal{D}}$,

$$\|\boldsymbol{d}\|_{\mathcal{D}} = \Big(\sum\nolimits_{i=1}^{q} d_i \overline{d_i} \Big)^{1/2}, \tag{9}$$

where d_i is the i^{th} component of $\boldsymbol{d}$ and $\bar{}$ denotes the complex conjugate.

2.3 Inversion via classical FWI

The essence of FWI ([61, 51, 67]) is to try to reconstruct the subsurface image by minimizing a misfit functional defined as the difference between the observations and simulations, starting from an initial model. In order to focus on reconstruction of deeper structures, we have to remove the direct arrivals from the forward model. Let us denote by $p_\omega^{(h)}$ and $p_{s,\omega}^{(h)}$ the solutions of (2) for m and m_s, where m_s is a smooth version of m. For the source h at frequency ω, we can define the forward operator at frequency ω for a source h by:

$$\begin{aligned} \mathcal{F}_\omega^{(h)}(m) &= p_\omega^{(h)}(\mathbf{x})|_\Sigma - p_{s,\omega}^{(h)}(\mathbf{x})|_\Sigma \\ &= \mathcal{R}\big(p_\omega^{(h)}(\mathbf{x})\big) - \mathcal{R}\big(p_{s,\omega}^{(h)}(\mathbf{x})\big), \end{aligned} \tag{10}$$

where $\mathcal{R}$ denotes the restriction operator to the discrete receivers location. When computing derivatives of $\mathcal{F}$, it will be necessary to remember that m_s depends also on m.

FWI amounts to solve the non linear least squares minimization problem:

$$\min_{m \in \mathcal{M}} \mathcal{J}(m) = \frac{1}{2} \sum_\omega \sum_h \|\mathcal{F}_\omega^{(h)}(m) - d_\omega^{(h)}\|_{\mathcal{D}}^2. \tag{11}$$

where the data $d_\omega^{(h)}$ are made of the observations at frequency ω for a source h, deprived from first arrivals. For sake of conciseness, we use the vector notation

$$\mathcal{F}(m) = \Big\{ \mathcal{F}_\omega^{(h)}(m), \text{for all } h, \omega \Big\} \quad \text{and} \quad \boldsymbol{d} = \Big\{ d_\omega^{(h)}, \text{for all } h, \omega \Big\}, \tag{12}$$

and write the FWI minimization problem as:

$$\min_{m \in \mathcal{M}} \mathcal{J}(m) = \frac{1}{2} \|\mathcal{F}(m) - \boldsymbol{d}\|_{\mathcal{D}}^2 \, . \tag{13}$$

As mentioned in the introduction, different norms can be used to build the objective function, and the minimization is usually performed by a Quasi-Newton algorithm, which requires only the gradient of the cost function:

$$\nabla \mathcal{J}(m) = D\mathcal{F}(m)^* \, (\mathcal{F}(m) - \boldsymbol{d}), \tag{14}$$

where $D\mathcal{F}$ stands for the Fréchet derivative of $\mathcal{F}$ and * is the adjoint. This gradient can be efficiently computed by the adjoint method, which does not require the formation of the Jacobian matrix. Appendix A describes a careful adaptation of this method to the case of complex variables (contrarily to the time-domain formulation, the data and the wavefields are, in the harmonic formulation, complex).

It is well known that the determination of the background (low spatial frequencies) of m by (13) is hampered by the many local minima of $\mathcal{J}$, caused by phase shifts in the synthetics (see Figure 18). This can be overcome only if the data contain extremely low frequencies. Appendix C shows a few reconstructions of velocity models via FWI, which illustrate the difficulties inherent to this approach. These difficulties are a motivation for alternative techniques such as the MBTT reformulation of FWI below, and for the optimizability study developed in this paper.

2.4 Inversion via MBTT–FWI (*background/data-space-reflectivity* decomposition)

In the MBTT (Migration-Based Traveltime) approach, see [24], the Earth model m is parameterized by a *smooth background* $\mathfrak{p}$ and a *data-space reflectivity* $\mathfrak{s}$ using a *migration* operator:

$$m = m(\mathfrak{p}, \mathfrak{s}) = \mathfrak{p} + \mathfrak{r} = \mathfrak{p} + \mathcal{W}\, D\mathcal{F}(\mathfrak{p})^*\, \mathfrak{s} = \mathfrak{p} + \sum_{\omega} \mathcal{W}(\omega) D\mathcal{F}_\omega(\mathfrak{p})^* \mathfrak{s}(\omega), \tag{15}$$

where $\mathfrak{r}$ is the depth reflectivity associated to $\mathfrak{s}$ and $\mathfrak{p}$; $\mathcal{W}$ is a scaling operator (which possibly depends on the frequency) and * denotes the adjoint. The weight $\mathcal{W}$ is meant to compensate for the lower amplitude of deep migrated events, see [46]. In our experiments we use a simple scaling proportional to the square root of the depth. where (x, y, z) are the 3D space coordinates with z the depth, and β is a scalar coefficient. We refer to Appendix B for more details regarding the computational aspects of the decomposition. Note that the solution of the linearized version (Born approximation) of the FWI problem (13) is of the form (15), in which case parameterization (15) is not *underparameterizing*. Hence, when the full model (2) is used, the parameterization by the data space reflectivity $\mathfrak{s}$ will be able to generate all primary events of the data, but maybe not all *multiple* events (i.e., it will miss the events associated to multiple reflections involving at least one reflector which generates no primary reflection), cf. [20].

When this decomposition is employed, the natural choice for the smooth version m_s of m in (10) (to account for the direct arrivals in the forward map $\mathcal{F}$) is simply $m_s = \mathfrak{p}$. With this change of parameter, the forward map given by (10) (12) rewrites:

$$\mathsf{F}(\mathfrak{p}, \mathfrak{s}) \stackrel{\text{def}}{=} \mathcal{F}(m) \qquad \text{with } \mathcal{F} \text{ given by (10) and } m \text{ by (15).} \tag{16}$$

By construction, $\mathcal{F}$ contains no direct arrivals, which implies that, for a smooth enough background $\mathfrak{p}$, F satisfies:

$$\mathsf{F}(\mathfrak{p}, 0) \stackrel{\text{def}}{=} \mathcal{F}(\mathfrak{p}) \approx 0. \tag{17}$$

The motivation for this parameterization is to eliminate phase shifts induced in the synthetics by changes in the background $\mathfrak{p}$: the events in the synthetic section $\mathsf{F}(\mathfrak{p}, \mathfrak{s})$ are obtained from the *data-space* reflectivity $\mathfrak{s}$ by migration followed by simulation with *the same kinematic*, and hence are expected to have the same phase as those of $\mathfrak{s}$, as illustrated in Figure 18, Section 5.1. Besides controlling the phase, this migration-demigration process has the additional property that the stack involved in the migration turns, for a fixed data space reflectivity, the data misfit into a coherency measure for the current background [20].

The MBTT–FWI minimization problem is then:

$$\min_{\mathfrak{p} \in \mathcal{M}_s, \mathfrak{s} \in \mathcal{D}} \mathsf{J}(\mathfrak{p}, \mathfrak{s}) = \frac{1}{2} \|\mathsf{F}(\mathfrak{p}, \mathfrak{s}) - \boldsymbol{d}\|_{\mathcal{D}}^2 \,, \tag{18}$$

where $\mathcal{M}_s \subset \mathcal{M}$ is the set of admissible smooth backgrounds and $\mathcal{D}$ is the data space.

The iterative reconstruction follows alternative updates of the two components of the model:

1. Start the reconstruction with the recovery of the reflectivity part $\mathfrak{s}$, from the lowest available frequency.

2. Using the updated reflectivity, which we fix, we can iterate to reconstruct the background part of the model, $\mathfrak{p}$.

3. Continue to perform the alternative updates, in a succession of increasing frequency.

This approach has been shown successful in [22], for the inversion of synthetic Marmousi data starting from a lower frequency of 5 Hz. Hence another motivation for the optimizability study of this paper is to quantify how far the MBTT reformulation of FWI succeeds in overcoming in step 2 the local minima problem inherent to classical FWI.

3 Optimizability and Attraction Basins

In this section, *optimizability* conditions are defined for general least-squares minimization problems, such as (13) and (18). By optimizability, we refer to the possibility for a local (deterministic) optimization algorithm to converge to a global minimum, i.e. there is no local minimum to hamper the search of the solution. This analysis follows the work of [19]. In addition to the notations of Subsection 2.2, we shall refer to $\mathcal{F}(\mathcal{M})$ as the *attainable set* of the least squares problem.

Definition 1 (Path)**.** *A curve P drawn on $\mathcal{F}(\mathcal{M}) \subset \mathcal{D}$ is a* path *of $\mathcal{F}(\mathcal{M})$ if it is of the form:*

$$P \ : \ t \in [0,1] \to \mathcal{F}\big((1-t)m_1 + tm_2\big) \qquad \text{where } m_1, m_2 \text{ are two models of } \mathcal{M}. \tag{19}$$

The following set of hypotheses is required to derive the optimizability conditions (cf. [19]):

- the forward map $\mathcal{F} : \mathcal{M} \to \mathcal{D}$ is twice differentiable along segments of $\mathcal{M}$,
- there exists $\mathcal{C} \geq 0$ such that $\forall m_1, m_2 \in \mathcal{M}, \forall t \in [0,1]$, $\|D_t\mathcal{F}\big((1-t)m_1+tm_2\big)\|_{\mathcal{D}} \leq \mathcal{C}\|m_2-m_1\|_E$, where D_t stands for the derivative with respect to t.

Note that $\mathcal{F}$ is indeed twice differentiable because our study is conducted in the finite (discrete) dimensional setting.

Definition 2 (Velocity and acceleration)**.** *P is twice differentiable and we denote by V and A (velocity and acceleration along P) its two first derivatives:*

$$V(t) = P'(t), \qquad A(t) = P''(t). \tag{20}$$

For simplicity, we shall consider only paths for which $V(t) \neq 0$ for all t, so we can define the unit tangent velocity v, and the normal acceleration a by

$$v(t) = \frac{V(t)}{\|V(t)\|_{\mathcal{D}}}, \qquad a(t) = \frac{A(t) - \langle A(t), v(t)\rangle\, v(t)}{\|V(t)\|^2_{\mathcal{D}}}, \tag{21}$$

where $\langle \cdot, \cdot \rangle$ is the real inner product in $\mathcal{D} = \mathbb{C}^q$:

$$\langle z_1, z_2 \rangle = Re\{z_1 \bar{z_2}\}. \tag{22}$$

Let us point out that, due to the limited accuracy of the recording devices, model error and noise, the observed data $\boldsymbol{d}$ do not belong in general to the attainable set $\mathcal{F}(\mathcal{M})$. Therefore, it is important that the least squares misfit function does not have parasitic local minima for data $\boldsymbol{d}$ which are "not too far" from the attainable set. This property is made precise by the following definition.

Definition 3 (Optimizability/Attraction Basin)**.** *The least squares problem* (13) *is* optimizable *on $\mathcal{M}$, or equivalently the* parameter set *$\mathcal{M}$ is an* attraction basin *for* (13)*, if there exists a neighborhood $\mathcal{V}$ of $\mathcal{F}(\mathcal{M})$ such that*

- *uniqueness: all data $\boldsymbol{d} \in \mathcal{V}$ have a unique projection $\boldsymbol{d}_\dagger$ on $\mathcal{F}(\mathcal{M})$,*
- *unimodality: for any $\boldsymbol{d} \in \mathcal{V}$, the distance to $\boldsymbol{d}$ has no parasitic stationary point over $\mathcal{F}(\mathcal{M})$,*
- *convergence: if $\boldsymbol{d} \in \mathcal{V}$, any minimizing sequence $d_n \in \mathcal{F}(\mathcal{M})$ of the distance to $\boldsymbol{d}$ is a Cauchy sequence for both the norm $\|\cdot\|_{\mathcal{D}}$ and the arc length distance $\ell(P)$ along the path P defined by* (19)*. Hence d_n converges in F to the unique projection $\boldsymbol{d}_\dagger$ of $\boldsymbol{d}$ onto $\mathcal{F}(\mathcal{M})$.*

Therefore, an *optimizable* least squares problem (13) exhibits rightful dispositions for its resolution by a local gradient algorithm: uniqueness of the projection and absence of local minimum ensure that the algorithm will converge to a global minimizer whatever the initial guess in its *basin of attraction* $\mathcal{M}$.

3.1 Global Radius of Curvature and Deflection

Following [19, pp. 167–172 and 300-308], we define the *global radius of curvature* and the *deflection* along a path P, and further give in Propositions 1 and 2 a characterization and a sufficient condition of optimizability.

Definition 4 (Radius of curvature)**.** *The (possibly infinite) radius of curvature $R(t)$ of a path P at t is given by:*

$$\frac{1}{R(t)} = \|a(t)\|_{\mathcal{D}} \;=\; \frac{\|A(t)\|_{\mathcal{D}}}{\|V(t)\|_{\mathcal{D}}^2}\sin(A(t),V(t)). \tag{23}$$

The radius of curvature of the whole path P is then defined as:

$$\frac{1}{R(P)} \stackrel{\text{def}}{=} \sup_{t\in[0,1]} \frac{1}{R(t)}. \tag{24}$$

It is straightforward to see that

$$\frac{1}{R(t)} \leq \frac{\|A(t)\|_{\mathcal{D}}}{\|V(t)\|_{\mathcal{D}}^2}, \qquad \text{for a.e. } t\in[0,1]. \tag{25}$$

Definition 5 (Global radius of curvature)**.** *The (possibly infinite)* global radius of curvature $R_G(t,t')$ *of a path P at t seen from t', with $t\neq t'$, is given by:*

$$R_G(t,t') = \left\{ \begin{array}{ll} N^+/D & \text{if } \langle v(t),v(t')\rangle \geq 0\ , \\ N^+ & \text{if } \langle v(t),v(t')\rangle \leq 0\ , \end{array} \right. \tag{26}$$

where $v(t), v(t')$ are the normalized velocities along P defined by (21)*, and*

$$\left\{ \begin{array}{lcl} N^+ & = & max(N,0) \quad \text{where } N = \text{sign}(t'-t)\langle P(t')-P(t), v(t')\rangle, \\ D & = & \Big(1-\langle v(t),v(t')\rangle^2\Big)^{1/2}. \end{array} \right. \tag{27}$$

The global radius of curvature of the path P is then defined as:

$$R_G(P) \stackrel{\text{def}}{=} \inf_{t,t'\in[0,1]} R_G(t,t') \geq 0. \tag{28}$$

The interest of global radius of curvature comes from the following proposition.

Proposition 1 ($R_G>0 \Longleftrightarrow$ optimizability)**.**
The least squares problem (13) *is* optimizable *- or equivalently $\mathcal{M}$ is an* attraction basin *for* (13) *- if and only if it exists $R_G>0$ such that $R_G(P)\geq R_G>0$ for all path P of $\mathcal{F}(\mathcal{M})$. The associated neighborhood $\mathcal{V}$ is defined by:*

$$\mathcal{V} = \big\{ \boldsymbol{d}\in F \quad | \quad \text{dist}\left(\boldsymbol{d}, \mathcal{F}(\mathcal{M})\right) < R_G \big\}. \tag{29}$$

The proofs can be found in [19]. The global radius of curvature can be computed numerically using (26) and (27), as will be done in following numerical sections. It can also be estimated via the usual radius of curvature depending on the value of the *deflection*, which we define now, and which is illustrated Figure 2(a).

Definition 6 (Deflection)**.** *The deflection between two points t and t' of the curve P is the angle between the two velocities $V(t)$ and $V(t')$ (see Figure 2(a)). It is given by:*

$$\Theta(t,t') = \arccos\left(\frac{\langle\, V(t), V(t')\,\rangle}{\|V(t)\|_{\mathcal{D}}\|V(t')\|_{\mathcal{D}}}\right) \in [0,\pi[\ . \tag{30}$$

The deflection $\Theta(P)$ of the curve P is defined as the largest angle $\Theta(t,t')\in[0,\pi]$ between any two tangent vectors $V(t)$ and $V(t')$ for any two points t and t' of $[0,1]$. An infinitesimal variation of the deflection $\mathrm{d}\Theta$ satisfies

$$\mathrm{d}\Theta \leq \frac{\|A(t)\|_{\mathcal{D}}}{\|V(t)\|_{\mathcal{D}}}\,\mathrm{d}t. \tag{31}$$

Denoting t_1 and t_2 the values of t for which the deflection is maximum, the deflection $\Theta(P)$ along the curve P satisfies

$$\Theta(P) = \int_{t_1}^{t_2} \mathrm{d}\Theta \leq \int_0^1 \frac{\|A(t)\|_{\mathcal{D}}}{\|V(t)\|_{\mathcal{D}}}\,\mathrm{d}t. \tag{32}$$

This majoration is sharp, but it is very conservative: equality holds only when P is an arc of circle with constant velocity $\|V(t)\|$: "when the path P turns always in the same direction with a constant radius". Therefore, the estimation of the deflection $\Theta(P)$ of the path using its upper bound in (32) is the *worst case estimate*.

The relation between global and local radii of curvature is then given by the following proposition.

Proposition 2 (Local and Global Radii of curvature)**.** *For any path P of $\mathcal{F}(\mathcal{M})$ one has*

$$R(P) \geq R_G(P) \geq 0 \quad \textit{and} \quad R(P) = R_G(P) \qquad \textit{as soon as} \quad \Theta(P) \leq \pi/2, \tag{33}$$

Definition 7 (Finite Curvature/Limited deflection (FC/LD) problem)**.** *The minimization Problem* (13) *is a FC/LD least-squares problem if:*

$$\textit{there exists } R > 0 \textit{ such that: } \quad \|A(t)\|_{\mathcal{D}} \leq \frac{1}{R}\|V(t)\|_{\mathcal{D}}^2 \quad \textit{for a.e. } t \in [0,1] \textit{ and all paths } P, \tag{34}$$

$$\Theta(P) \leq \frac{\pi}{2} \quad \textit{for all paths } P. \tag{35}$$

From Definition 7, a FC/LD problem verifies that, using (33):

$$R_G(P) = R(P) \geq R > 0 \quad \text{for all paths } P, \tag{36}$$

which shows that FC/LD problems (also referred to as *weakly nonlinear* inverse problem in [23]) are necessarily optimizable.

Notice that Proposition 1 gives a characterization of optimizable problems, whereas Definition 7 provides only a sufficient condition.

3.2 Directional Attraction Basins

Numerical application of previous section to check whether or not a given least squares problem is optimizable becomes quickly intractable when the number of parameters n increases, as it is the case in seismic inversion. So we limit ourselves to *directional* (or *one-dimensional*) *parameter sets* of the form:

$$\mathcal{M}(m_0, \mathfrak{u}, \Delta) = [m_0 - \Delta\mathfrak{u}, m_0 + \Delta\mathfrak{u}] \quad \text{with} \quad \|\mathfrak{u}\|_E = 1. \tag{37}$$

Here, m_0 is a nominal model, $\mathfrak{u}$ a normalized perturbation direction, and Δ gives the size of the domain of investigation. The associated attainable set is constructed from the path P defined by:

$$P \;:\quad t \in [0,1] \rightsquigarrow \mathcal{F}\big(m_0 + (2t-1)\Delta\mathfrak{u}\big). \tag{38}$$

We refer to directional optimizability when the problem is optimizable for an interval such as (37), and this interval is a directional attraction basin. Directional optimizability is only a necessary condition for optimizability, but it will allow to analyze the behavior of seismic inverse problems and to compare formulations: the size Δ of a directional attraction basin in a descent direction tells us how far one can move away in this direction without being stopped by parasitic local minima. Our objective now is to determine (see illustration Figure 2):

1. the *size* $\Delta_{m_0}^{\mathfrak{u}}$ of the directional attraction basin centered at m_0. The larger $\Delta_{m_0}^{\mathfrak{u}}$ is, the better the least squares problem is amenable to minimization by local algorithm, because we allow a larger area for investigation. In our numerical experiments, we shall scale the estimate with the norm of the nominal model, $\|m_0\|_E$, to provide relative (to the model) quantity.

2. the associated *tolerable error level* $R^{\mathfrak{u}}_{G,m_0}$. It is the largest tolerable error on the data $\boldsymbol{d}$ which ensures the absence of parasitic local minima for the least squares objective function

$$t \in [0,1] \rightsquigarrow \frac{1}{2}\|\mathcal{F}(m_0 + (2t-1)\Delta^{\mathfrak{u}}_{m_0}\mathfrak{u}) - \boldsymbol{d}\|^2_{\mathcal{D}} \tag{39}$$

over $[0,1]$. The larger $R^{\mathfrak{u}}_{G,m_0}$ is, the better is the robustness of the minimization procedure to noise in the data. In our numerical experiments, we divide the estimates with the norm of the synthetic data $\boldsymbol{d}_0 = \mathcal{F}(m_0)$ to provide relative (to the data) quantity.

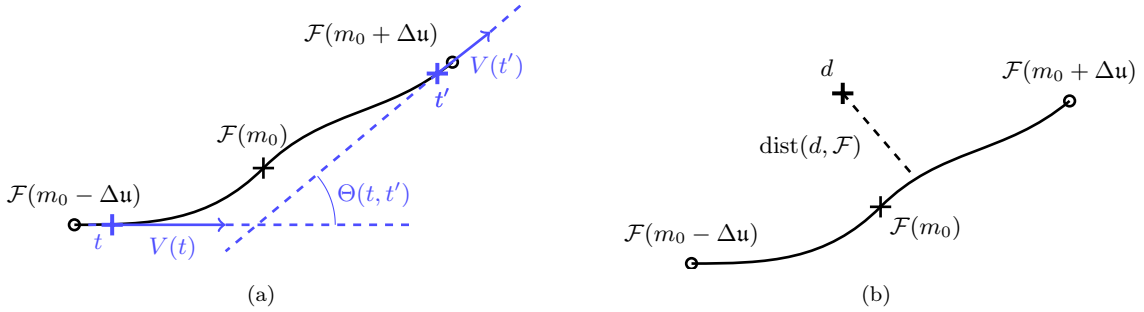


Figure 2: A one-dimensional setup for least squares problems. The figure lives in the data space, the attainable set is the path P image of the interval $\mathcal{M}(m_0, \mathfrak{u}, \Delta)$ of the model space. (a) Illustration of the computation of the deflection $\Theta(t,t')$ between two arbitrary points t and t'. (b) The path has a finite curvature and the deflection is smaller than $\pi/2$, so the FC/LD Property 7 is satisfied, and $R_G = R > 0$ by Proposition 2. Hence the "distance to $\boldsymbol{d}$" function cannot have local minimum over P provided the data $\boldsymbol{d}$ is at a distance of the attainable set $P = \mathcal{F}(\mathcal{M}(m_0, \mathfrak{u}, \Delta))$ smaller than R.

We shall use two types of estimate:

- Θ-estimates of $\Delta^{\mathfrak{u}}_{m_0}$, where the optimizability over $\mathcal{M}$ is obtained by satisfying the sufficient condition $\Theta(P) \leq \pi/2$ of Definition 7. In this case, $R_G(P) = R(P)$, so the tolerable error level is given by the minimum over the $[0,1]$ interval of $R(t)$ given by (23).
- R_G-estimates of $\Delta^{\mathfrak{u}}_{m_0}$, where optimizability is obtained by satisfying the $R_G > 0$ characterization of optimizability of Proposition 1. In this case, the associated tolerable error level R_G has to be computed by evaluating numerically the infimum in (28) using (26) and (27).

Remark 2. *The size of attraction basins depends only on the choice of the forward map to be inverted, but not on the optimization algorithm used (e.g. Newton method, gradient descent, etc). The choice of the method naturally affects the rate of convergence and the speed at which the final solution is eventually reached, but it has no influence on the presence or absence of local minimum.*

3.3 Local Θ-estimate for $\Delta^{\mathfrak{u}}_{m_0}$ and associated tolerable error level $R^{\mathfrak{u}}_{m_0}$

We provide here a *local* Θ*-estimate* Δ of the attraction basin, in the sense that it is *based only on* the velocity V and acceleration A *at* m_0 *in the direction* $\mathfrak{u}$. In order to ensure that the deflection of the path P defined by (38) is smaller than $\pi/2$, we use the upper bound (32) of $\Theta(P)$, which, according to the optimizability condition of definition 7, ensures in turn that $\mathcal{M}(m_0, \mathfrak{u}, \Delta)$ is an attraction basin.

With the notations of [17] for the directional derivative, the chain rule differentiation gives:

$$\begin{cases} V(t) = \dfrac{\partial \mathcal{F}}{\partial m}\dfrac{\partial m}{\partial t} = 2\Delta D\mathcal{F}(m)(\mathfrak{u}), \\ A(t) = \dfrac{\partial^2 \mathcal{F}}{\partial m^2}\left(\dfrac{\partial m}{\partial t}\right)^2 + \dfrac{\partial \mathcal{F}}{\partial m}\dfrac{\partial^2 m}{\partial t^2} = 4\Delta^2 D^2\mathcal{F}(m)(\mathfrak{u},\mathfrak{u}), \end{cases} \tag{40}$$

where $\mathfrak{u}$ acts as the direction of derivation. Then we use a rectangle approximation in equation (32), which gives the approximate *upper bound* $\Theta^{\mathfrak{u}}_{m_0}$ to the deflection $\Theta(P)$:

$$\Theta(P) \leq \int_0^1 \frac{\|A(t)\|_{\mathcal{D}}}{\|V(t)\|_{\mathcal{D}}}\,\mathrm{d}t \sim \frac{\|A(1/2)\|_{\mathcal{D}}}{\|V(1/2)\|_{\mathcal{D}}} = 2\Delta\frac{\|D^2\mathcal{F}(m_0)(\mathfrak{u},\mathfrak{u})\|_{\mathcal{D}}}{\|D\mathcal{F}(m_0)(\mathfrak{u})\|_{\mathcal{D}}} \overset{\text{def}}{=} \Theta^{\mathfrak{u}}_{m_0}. \tag{41}$$

This gives immediately a *local* Θ*-estimate* of the size Δ of an attraction basin at m_0 in the direction $\mathfrak{u}$:

$$\Delta^{\mathfrak{u}}_{m_0} = \frac{\pi}{4}\frac{\|D\mathcal{F}(m_0)(\mathfrak{u})\|_{\mathcal{D}}}{\|D^2\mathcal{F}(m_0)(\mathfrak{u},\mathfrak{u})\|_{\mathcal{D}}}, \qquad \text{local } \Theta\text{-estimate.} \tag{42}$$

This estimate is an *approximate* (because of the rectangle approximation of the integral) *lower bound* (because it is based on the upper bound (32)) to the size of the largest attraction basins at m_0 in the direction $\mathfrak{u}$. It is computationally cheap, and we will use it in Section 4 to analyze the effect of frequency and perturbation direction on the optimizability of the FWI least squares problem.

The associated *tolerable error level* $R^{\mathfrak{u}}_{m_0}$ is then the minimum of the radius of curvature along $P = \mathcal{F}(\mathcal{M}(m_0,\mathfrak{u},\Delta))$, which is approximated simply by its value at m_0, that is $R(t = 1/2)$ given by (23):

$$R^{\mathfrak{u}}_{m_0} = \left(\frac{\|V(t)\|^2_{\mathcal{D}}}{\|A(t)\|_{\mathcal{D}}|\sin\big(A(t),V(t)\big)|}\right)\Big|_{t=\frac{1}{2}}, \qquad \text{tolerable error level,} \tag{43}$$

where $V(t)$ and $A(t)$ have been defined in (40).

Remark 3. *One may use the directional attraction basins* $\mathcal{M}(m_0,\mathfrak{u},\Delta^{\mathfrak{u}}_{m_0})$ *to construct an approximate multidimensional attraction basin* $\mathcal{M}(m_0,\Delta_{m_0})$ *as follows: suppose that* $\Delta^{\mathfrak{u}}_{m_0} \geq \Delta_{m_0} \geq 0$ *for all* $\mathfrak{u}$ *for some* Δ_{m_0}*, and define:*

$$\mathcal{M}(m_0,\Delta_{m_0}) = B(m_0,\Delta_{m_0}), \tag{44}$$

where B *denotes a ball in the model space. The size* Δ_{m_0} *of an approximate multidimensional attraction basin* $\mathcal{M}(m_0,\Delta_{m_0})$ *follows immediately from the previous results by computing an upper bound* Θ_{m_0} *to* $\Theta^{\mathfrak{u}}_{m_0}$ *using* (41)*:*

$$\Theta^{\mathfrak{u}}_{m_0} = 2\Delta\frac{\|D^2\mathcal{F}(m_0)(\mathfrak{u},\mathfrak{u})\|}{\|D\mathcal{F}(m_0)(\mathfrak{u})\|} \leq 2\Delta\frac{\|D^2\mathcal{F}(m_0)\|\|\mathfrak{u}\|^2}{\lambda^{D\mathcal{F}}_{\min}\|\mathfrak{u}\|} = 2\Delta\frac{\|D^2\mathcal{F}(m_0)\|}{\lambda^{D\mathcal{F}}_{\min}} \stackrel{\text{def}}{=} \Theta_{m_0}\ , \tag{45}$$

where $\lambda^{D\mathcal{F}}_{\min}$ *denotes the lowest singular value of* $D\mathcal{F}(m_0)$*, and we use the Frobenius norm for the matrix norm of the numerator. Estimation of* Δ_{m_0} *follows:*

$$\Delta_{m_0} = \frac{\pi}{4}\frac{\|\lambda^{D\mathcal{F}}_{\min}\|}{\|D^2\mathcal{F}(m_0)\|}\ . \tag{46}$$

Note that $\mathcal{M}(m_0,\Delta_{m_0})$ *is only an* approximate *attraction basin, because condition* $R_G > 0$ *(Proposition 1) or* $\Theta \leq \pi/2$ *(Definition 7) holds only for segments passing through its center* m_0*.*

Then, (43) *provides a lower bound* R_{m_0} *independent of* $\mathfrak{u}$ *to the radius of curvature of* $\mathcal{M}(m_0,\Delta)$*:*

$$R_{m_0} = \frac{\big(\lambda^{D\mathcal{F}(m_0)}_{\min}\big)^2}{\|D^2\mathcal{F}(m_0)\|}\ . \tag{47}$$

3.4 Exact Θ- and R_G-estimates of $\Delta^{\mathfrak{u}}_{m_0}$ and associated tolerable error levels

The determination of *the exact* Θ*- and* R_G*-estimates* of the attraction basin centered at m_0 in a direction $\mathfrak{u}$ inside an interval $\mathcal{M}(m_0,\mathfrak{u},\Delta)$ of given size Δ requires the numerical computation of the deflection $\Theta(t,t')$ and the global radius of curvature $R_G(t,t')$ between any two points

$$\mathcal{F}(m_0+t\mathfrak{u})\ ,\ -\Delta \leq t \leq \Delta \quad \text{and} \quad \mathcal{F}(m_0+t'\mathfrak{u})\ ,\ -\Delta \leq t' \leq \Delta \tag{48}$$

of the path P, which is the image by $\mathcal{F}$ of the investigated interval $\mathcal{M}(m_0,\mathfrak{u},\Delta)$.

For this purpose,*deflection maps* and *global radius maps* are computed, and they display the values of $\Theta(t,t')$ (Definition 6) and of $R_G(t,t')$ (Definition 5) between the points of $\mathcal{M}(m_0,\mathfrak{u},\Delta)$. On the diagonal of the maps, where $t = t'$, R_G is not defined by (26) (27), and we indicate instead the values of $R(t)$ given by (21) (23), which represent the limits of $R_G(t,t')$ when $t' \to t$. One can then read on these maps the exact Θ-estimate (resp. the exact R_G-estimate) of $\Delta^{\mathfrak{u}}_{m_0}$ by searching for the largest squares centered at (0,0) where $\Theta(t,t') \leq \pi/2$ (resp. where $R_G(t,t') > 0$).

During this process, when Δ increases from 0 to its exact Θ-estimate, the associated exact tolerable error $R^{\mathfrak{u}}_{m_0} = \inf_{-\Delta\leq t\leq\Delta} R(t)$ decreases from its value R_0 at m_0 to the tolerable error $R^{\mathfrak{u}}_{m_0}$ of the Θ-attraction basin. When Δ increases further to its exact R_G-estimate, the tolerable error is $R_G = \inf_{-\Delta\leq t,t'\leq\Delta} R_G(t,t')$, which continues to decrease, until it reaches the value 0 of the R_G-attraction basin.

These computationally intensive *exact* estimates will be used in Section 4 on the analysis of FWI, for comparison purpose with the local Θ-estimates. They will also be used in Section 5 to compare more precisely the optimizability properties of classical FWI and MBTT–FWI.

4 FWI: Numerical determination of convergence basins and tolerable error levels

Our objective in this section is to give a comprehensive understanding of the behavior of the seismic FWI problem (13) with respect to parameters such as the frequency, the geometry of the target, the parametrization, etc. Therefore, we select a nominal model m_0 and compute, according to Subsection 3.2, the size $\Delta^{\mathfrak{u}}_{m_0}$ of an attraction basin and the associated tolerable error level $R^{\mathfrak{u}}_{m_0}$ for different (normalized) directions $\mathfrak{u}$. When m_0 is smooth, the direction $\mathfrak{u}$ can be seen as the search direction, or the geometry of the unknown parameter to be reconstructed in the framework of inversion.

4.1 Influence of frequency ω and search direction $\mathfrak{u}$

We consider the two-dimensional geophysical setup of Figure 1 for the acoustic Helmholtz equation (2), and select the squared slowness $m := c^{-2}$ as unknown parameter, which is expressed in $\mathrm{s}^2\,\mathrm{m}^{-2}$. The forward map $\mathcal{F}_\omega$ at frequency ω for parameter m is a vector of $\mathbb{C}^{(n_{rcv}\times n_{src})}$ with n_{rcv} the number of receivers per source and n_{src} the number of sources. In this experiment, we take 19 sources and 183 receivers associated with each source. Both are located near the surface (accordingly to Figure 1), note that the receivers remain in the same position for all sources. Thus, we work with reflection data obtained from a one side (the surface) illumination.

The nominal model m_0 is the smooth 9.2×3 km background velocity pictured in Figure 3, where the velocity varies linearly with depth, according to the fact that no information is known on the structures to reconstruct.

We investigate the behavior of attraction basins for different directions $\mathfrak{u}$, which are chosen to be based on straight reflectors, or extracted from the Marmousi model (probably the most popular velocity model in geophysics, synthetically designed by the Institut Français du Pétrole in 1988). They are shown in Figure 4: $\mathfrak{u}_1$ has a single, slim reflector (Figure 4(a)), $\mathfrak{u}_2$ has two slim reflectors (Figure 4(b)), $\mathfrak{u}_3$ has a single large reflector (Figure 4(c)) and $\mathfrak{u}_m$ is obtained by high-pass filtering from the Marmousi model of Figure 4(d). The amplitude for the directions are selected such that $\mathfrak{u}_k$ is normalized: $\|\mathfrak{u}_k\| = 1$, for all k.

Given a perturbation size Δ, one can think of $m_1 = m_0+\Delta\mathfrak{u}$ as being the target velocity, which one tries to retrieve starting from the smooth model m_0 by minimization of the FWI objective function in direction $\mathfrak{u}$ by a local algorithm (it is analogous to the update formula in Newton-type algorithm). The question is: how large can Δ be while still guaranteeing that the algorithm will not stop before the global minimum, provided that the error in the data is smaller than some R or R_G?

Local Θ-estimates of $\Delta^{\mathfrak{u}}_{m_0}$ and $R^{\mathfrak{u}}_{m_0}$

The determination of $\Delta^{\mathfrak{u}}_{m_0}$ with the local Θ-estimate, (42), and $R^{\mathfrak{u}}_{m_0}$ with (43), requires the computation of the first and second order directional derivatives in the direction $\mathfrak{u}$, and we refer to Appendix A.3 for the computational details. Figure 5 shows the evolution of $\Delta^{\mathfrak{u}}_{m_0}$ (top) and $R^{\mathfrak{u}}_{m_0}$ (bottom) for frequencies between 0.5 to 15 Hz, for the four directions $\mathfrak{u}$ of Figure 4.

We observe the following.

– In all cases, the size $\Delta^{\mathfrak{u}}_{m_0}$ of the directional attraction basin decreases with increasing frequency, meaning that the largest basin of attraction is obtained for the lowest frequencies. This is the

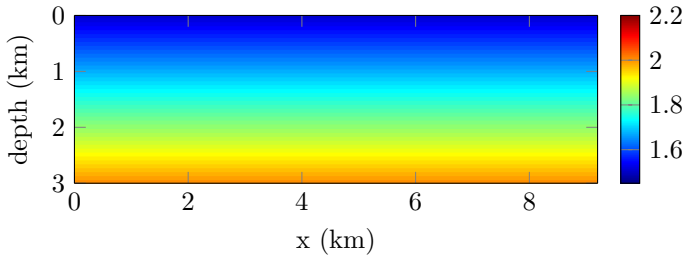
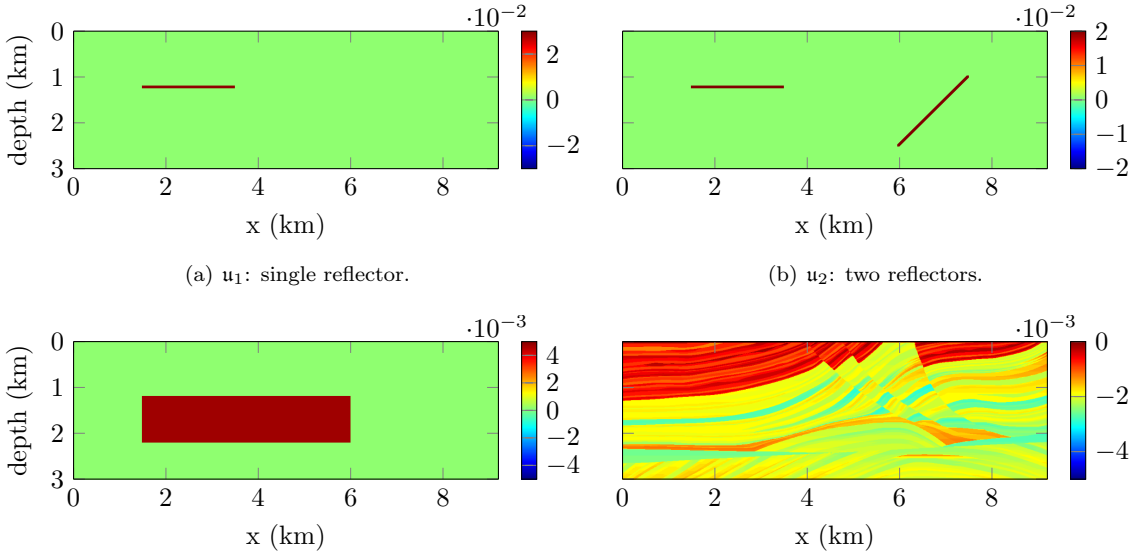


Figure 3: Acoustic model c_0 of size 9.2×3 km, the velocity is indicated in $\mathrm{km\,s^{-1}}$. The estimates are calculated at the nominal model $m_0 = c_0^{-2}$.



(a) $\mathfrak{u}_1$: single reflector.

(b) $\mathfrak{u}_2$: two reflectors.

(c) $\mathfrak{u}_3$: single large reflector.

(d) $\mathfrak{u}_m$: reflectors following Marmousi structure.

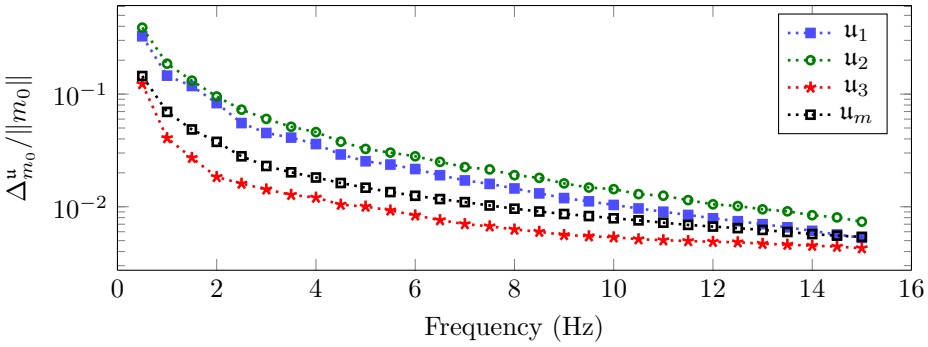
Figure 4: Four different directions for the estimation of the quantities of interest ($R^{\mathfrak{u}}_{m_0}$ and $\Delta^{\mathfrak{u}}_{m_0}$). The amplitude is determined such that $\|\mathfrak{u}_k\| = 1$. The values of $\mathfrak{u}_k$ are given in $(\mathrm{m\,s^{-1}})^{-2} = \mathrm{s^2\,m^{-2}}$.

expected behaviour that indicates that lower frequencies give a better chance for the iterative minimization to converge, especially when no prior information is known for the initial model.

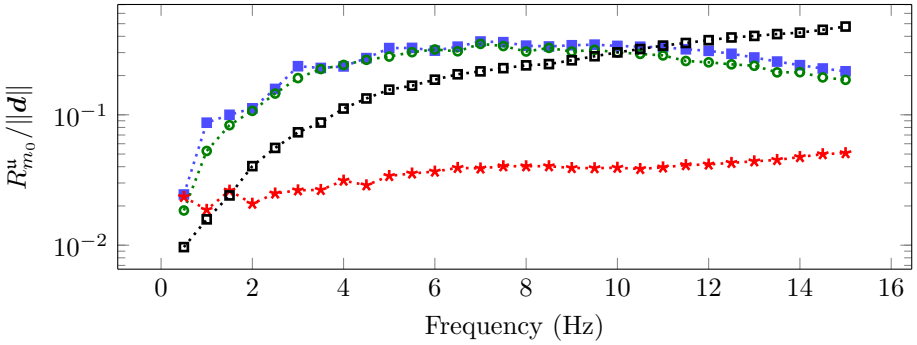
– Concerning the effect of the perturbation direction on attraction basins, Figure 5(a) shows that their size decreases in the order $\mathfrak{u}_1 \to \mathfrak{u}_2 \to \mathfrak{u}_m \to \mathfrak{u}_3$. This can be understood as follows: Directions $\mathfrak{u}_1$, $\mathfrak{u}_2$ and $\mathfrak{u}_m$ are essentially reflective, and contain no low frequency components. Hence increasing Δ will not change the phase of backscattered data, but because of multiple reflections some nonlinearity will occur, which limits the size of the attraction basins. The directions $\mathfrak{u}_1$ and $\mathfrak{u}_2$ with the less multiple reflections have the largest basins of attraction. Then comes $\mathfrak{u}_m$ which has many reflectors and hence many multiples. On the contrary, direction $\mathfrak{u}_3$ has a strong low frequency component besides its reflectivity. Hence increasing Δ will generate a phase shift in the backscattered data, which is likely to create local minima more quickly, and one expects the size $\Delta^{\mathfrak{u}}_{m_0}$ of the attraction basin in the direction $\mathfrak{u}_3$ to be much smaller than in the three other directions.

– Concerning the largest tolerable error, Figure 5(b) shows that $R^{\mathfrak{u}}_{m_0}$ increases with the frequency, meaning that the low frequencies are more likely to be affected in the presence of noise. We notice the exception of $\mathfrak{u}_3$ for which the allowed error remains relatively similar (and worse than the others) with frequency.

These results clearly show that it is harder to converge towards salt dome than it is to recover



(a) Evolution of $\Delta^{\mathfrak{u}}_{m_0}$ computed with local Θ-estimate.



(b) Evolution of the maximal distance between the data and the attainable set, $R^{\mathfrak{u}}_{m_0}$.

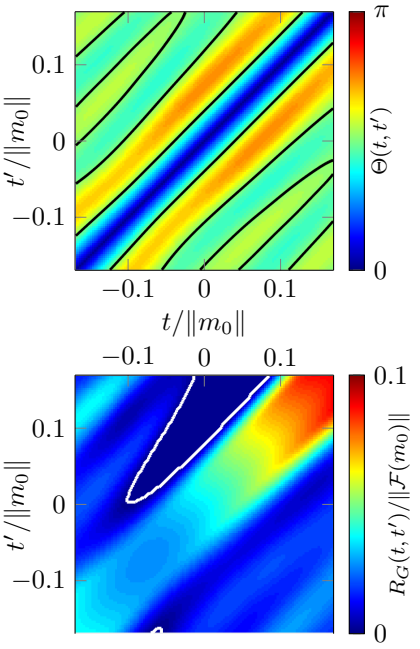
Figure 5: Evolution of $\Delta^{\mathfrak{u}}_{m_0}$ and $R^{\mathfrak{u}}_{m_0}$ with frequency using m_0 as a smooth velocity background of Figure 3, and using different directions $\mathfrak{u}$. The blue squares (··■··) employ one reflector ($\mathfrak{u}_1$, Figure 4(a)), the green circles (··●··) two reflectors ($\mathfrak{u}_2$, Figure 4(b)) the red stars (··★··) one large reflector ($\mathfrak{u}_3$, Figure 4(c)) and the black squares (··■··) the reflectors extracted from the Marmousi model ($\mathfrak{u}_m$, Figure 4(d)).

Marmousi-like structures. Because of its low spatial frequency component, the former reduces the size of the attraction basin and is less tolerant to noise in the data. This situation is further illustrated in the context of model reconstruction in Appendix C. Note that the order of magnitude in Figure 5 are around 10^{-1} of the norms of the model (top) and data (bottom).

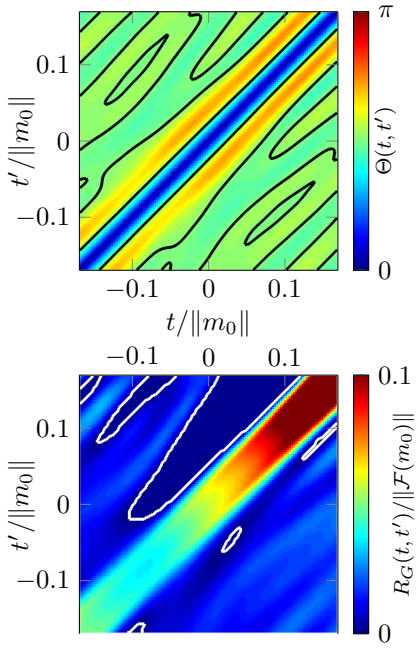
Exact Θ- and R_G-estimates of $\Delta^{\mathfrak{u}}_{m_0}$ and $R^{\mathfrak{u}}_{m_0}$

As an alternative to confirm these preliminary observations, we compute $\Delta^{\mathfrak{u}}_{m_0}$ using the exact Θ-estimate and R_G-estimate. It requires the computation of deflection and global radius maps on the investigated interval, from which the estimate is extracted, cf. Subsection 3.4. In Figures 6 and 7, we show the resulting maps in the direction of the Marmousi reflectors $\mathfrak{u}_m$ of Figure 4(d) and of the single large reflector $\mathfrak{u}_3$ of Figure 4(c), using frequencies 4 and 7 Hz. In Table 1, we show the numerical values of the interval size centered around m_0, depending on the estimation.

The maps confirm the results of the previous experiment: the exact Θ-estimates of $\Delta^{\mathfrak{u}}_{m_0}$ (top of the figures) are of the same order of magnitude, but slightly larger than the local Θ-estimates of Figure 5, cf. Table 1. Similarly, the exact R_G-estimates of $\Delta^{\mathfrak{u}}_{m_0}$ (bottom of the figures), are larger (Table 1), but at the price of a smaller tolerable error on the data. We also note a much more consistent pattern in the deflection for $\mathfrak{u}_m$, compared to $\mathfrak{u}_3$. Comparing the formulas, we observe that the inexpensive local Θ-estimates already provide accurate values for classical FWI problems, close to the more refined (and computationally more intensive) exact Θ- and R_G-estimates. Therefore, we should only use the local Θ-estimate in the following of this section.



(a) Using the Marmousi direction $\mathfrak{u}_m$ and frequency 4Hz.



(b) Using the Marmousi direction $\mathfrak{u}_m$ and frequency 7Hz.

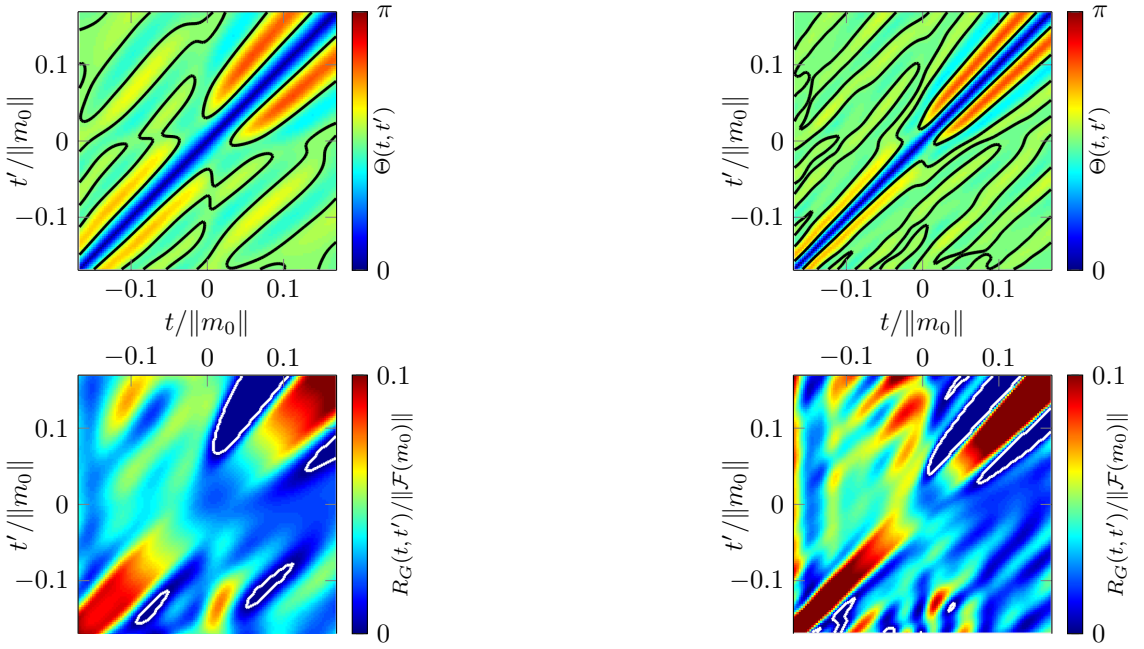
Figure 6: Maps of the deflection (30) and global radius (26) between two models of $\mathcal{M}(m_0, \mathfrak{u}_m, \Delta)$, see (48). The selected direction $\mathfrak{u}_m$ is Figure 4(d), m_0 and Figure 3. The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

	direction $\mathfrak{u}_m$		direction $\mathfrak{u}_3$	
	$\frac{\Delta^{\mathfrak{u}_m}_{m_0}}{\|m_0\|}$	$\frac{R^{\mathfrak{u}_m}_{G,m_0}}{\|\mathcal{F}(m_0)\|}$	$\frac{\Delta^{\mathfrak{u}_3}_{m_0}}{\|m_0\|}$	$\frac{R^{\mathfrak{u}_3}_{G,m_0}}{\|\mathcal{F}(m_0)\|}$
4 Hz				
Local Θ-estimates	0.018	0.10	0.012	0.03
Exact Θ-estimates	0.002	0.03	0.02	0.018
Exact R_G-estimates	0.003	0	0.05	0
7 Hz				
Local Θ-estimates	0.01	0.15	0.008	0.03
Exact Θ-estimates	0.01	0.05	0.01	0.016
Exact R_G-estimates	0.03	0	0.03	0

Table 1: Size Δ of attraction basins centered at m_0 using directions $\mathfrak{u}_m$ and $\mathfrak{u}_3$ (see Figure 4) and corresponding maximal tolerable error R_G, at 4 and 7 Hz. By construction, the R_G-estimates correspond to the limit case of a zero tolerable error, the other values are extracted from Figures 5, 6 and 7.

4.2 Influence of the initial model

We now investigate the influence of the initial model m_0 on the size of attraction basins. We consider two initial models: the acoustic Marmousi medium m_m and a model m_c encompassing objects of high contrasts, given in Figure 8. The local Θ-estimate of $\Delta^{\mathfrak{u}}_{m_j}$ given by (42) is plotted in Figure 9, for the two directions $\mathfrak{u}_1$ (one reflector) and $\mathfrak{u}_3$ (one large reflector). We see that $\Delta^{\mathfrak{u}}_{m_j}$ is barely affected by the choice of initial model for the computation. We still observe the decrease in the size of the interval with increasing frequency, and the deterioration in the direction of the large reflector $\mathfrak{u}_3$. Clearly, when comparing amplitudes of our estimates, the direction $\mathfrak{u}$ has much more influence than

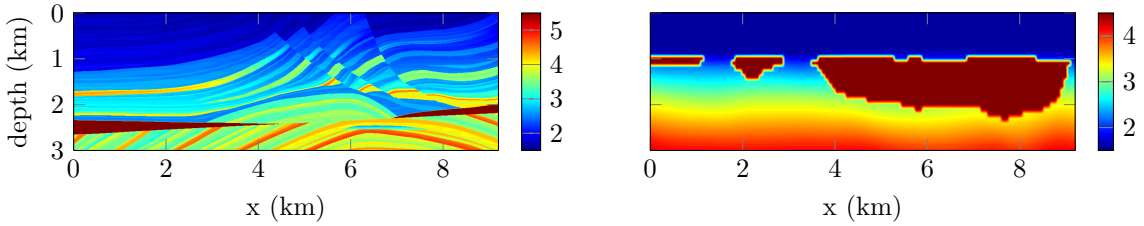


(a) Using the direction $\mathfrak{u}_3$ and frequency 4Hz.

(b) Using the direction $\mathfrak{u}_3$ and frequency 7Hz.

Figure 7: Maps of the deflection (30) and global radius (26) of $\mathcal{M}(m_0, \mathfrak{u}_3, \Delta)$, see (48). The selected direction $\mathfrak{u}_3$ is Figure 4(c), m_0 and Figure 3. The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

the initial model m_0.



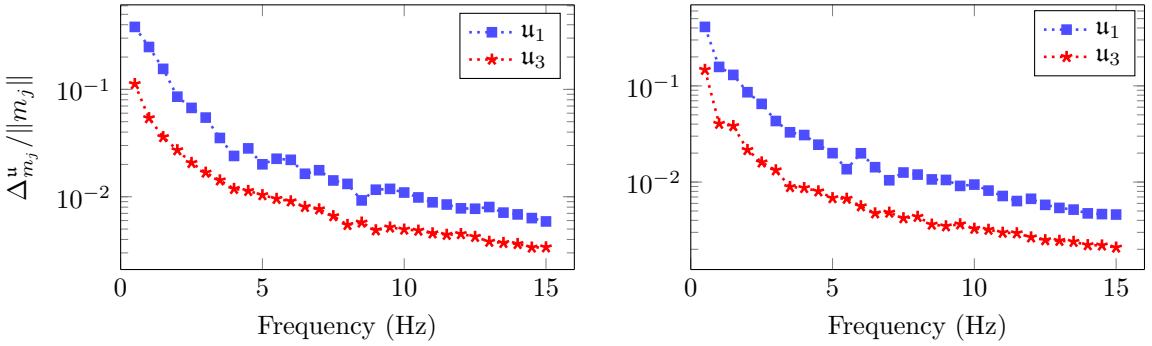
(a) Marmousi velocity model c_m.

(b) Velocity including high contrast objects c_c.

Figure 8: Acoustic models of size 9.2 × 3 km used as initial models for the estimation of $\Delta^{\mathfrak{u}}_{m_j}$, we take $m_j = c_j^{-2}$, for $j = \{m, c\}$. The velocity is indicated in $\mathrm{km\,s^{-1}}$.

4.3 Approximate n-dimensional attraction basins

We compare the directional local Θ-estimates of Subsection 4.1, evaluated at the smooth model m_0 of Figure 3 in the Marmousi direction $\mathfrak{u}_m$ of Figure 4(d), with the approximate multidimensional attraction basins centered in m_0 introduced in Remark 3. The determination of the radius Δ_{m_0} of such a spherical attraction basin by (46) and of the associated level of tolerable error R_{m_0} by (47) requires the computation of the lowest singular value of the Jacobian of $\mathcal{F}$, and the computation of its full second order derivative matrix. In order to reduce the computational burden associated with this computation, we develop a simple model reduction process where the wave speed in adjacent cells is averaged to generate a coarse representation. Note that this approach is also employed to improve the stability of the inverse problem, cf. [11]. When applied to the smooth acoustic model of



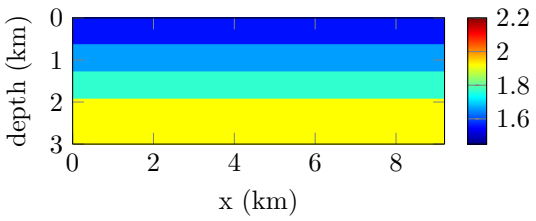
(a) Estimates using m_m of Figure 8(a).

(b) Estimates using m_c of Figure 8(b).

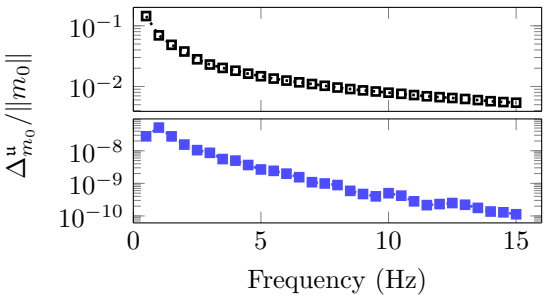
Figure 9: Evolution with frequency of $\Delta^{\mathfrak{u}}_m$ from local Θ-estimate for two initial models m_m (left) and m_c (right), in the direction of $\mathfrak{u}_1$ and $\mathfrak{u}_3$ (Figure 4).

Figure 3, this leads to the model of Figure 10(a), where the number of coefficients has been reduced to 56. Using this reduced representation, we can compute all quantities to effectively obtain the lower bounds Δ_{m_0} and R_{m_0} (i.e., we compute the matrix $D^2\mathcal{F}$ and the singular values of $D\mathcal{F}$).

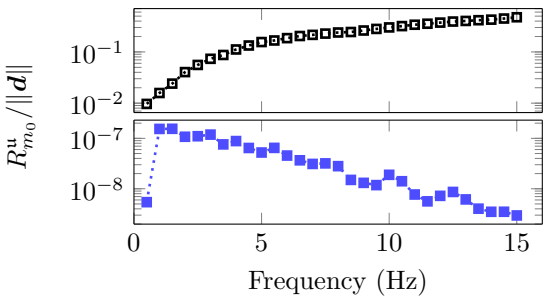
Figure 10(b) shows an important reduction of the size Δ_{m_0} of the n-dimensional basin compared to the directional estimates in the Marmousi direction u_m using the local Θ-estimate (six orders of magnitude). Yet, it confirms the main property, which is the decrease of the attraction basin size with increasing frequencies. A similar observation holds for the tolerable data error level R_{m_0}, see Figure 10(c), which suffers an important reduction compared to the directional estimates. In addition, we observe a different pattern between R_{m_0}, which decreases with frequency, and the directional estimates $R^{\mathfrak{u}}_{m_0}$, which increases with frequency. As illustrated in Figure 5, $R^{\mathfrak{u}}_{m_0}$ depends more on the direction than $\Delta^{\mathfrak{u}}_{m_0}$, and it is reflected on the n-dimensional estimates, which takes the worst case.



(a) Compression of the acoustic starting model of size 9.2×3 km, where only 56 coefficients are kept to represent the velocity, indicated in km s^{-1}.



(b) Evolution of $\Delta^{\mathfrak{u}}_{m_0}$ computed with local Θ-estimate.



(c) Maximal distance between the data and the attainable set.

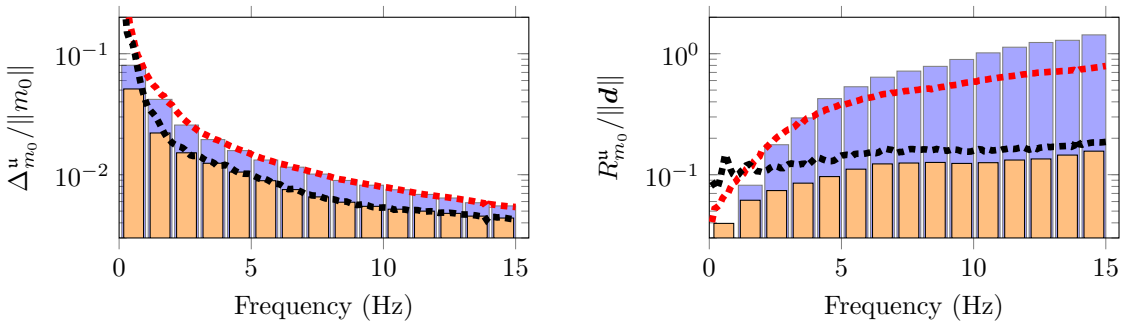
Figure 10: Lower bounds and directional estimates of $\Delta^{\mathfrak{u}}_{m_0}$ and $R^{\mathfrak{u}}_{m_0}$ using a model with reduced dimensions. The respective lower bounds are given with the blue squares (··■··), the black squares (··■··) show the directional estimates using $\mathfrak{u}_m$ (reflectors from the Marmousi model, see Figure 4(d)).

From this first set of experiments, we clearly see that the frequency should evolve from low to high values in order to facilitate the convergence. When no prior information is known for the initial model, one should take advantage of the larger interval size $\Delta^{\mathfrak{u}}_{m_0}$ given by low frequencies. Here we motivate the well-known frequency progression (e.g., [16, 59]) from quantitative estimates of the size of the basin of attraction, which allows the quantification of the benefits. However, the low frequencies require more accuracy regarding the data, as illustrated with the estimates of $R^{\mathfrak{u}}_{m_0}$. In addition, we mention the fact that seismic data are particularly affected by noise at low frequencies, which are sometimes unusable. This highlights where resides the difficulty: the low frequencies that are required to converge may be unusable due to the noise, or the inaccuracy may violate the curvature condition.

4.4 Frequency bandwidth data

We have seen that the sequential progression in frequency must be conducted from low to high regime. We now investigate the case of frequency bandwidth: when the forward problem encompasses not only one, but several frequencies. This approach is particularly natural for the time domain inverse problem where several frequencies are simultaneously contained in the data. The forward problem is now a vector of $\mathbb{C}^{(n_{rcv}\times n_{src}\times n_\omega)}$, where n_ω is the number of frequencies taken in the subgroup.

We remain with the smooth background of Figure 3 for m_0. We select groups of ten consecutive frequencies. The initial group encompasses the frequencies from 0.1 to 1 Hz using 0.1 Hz increment. Similarly, we design fifteen groups so that the last one is from 14.1 to 15 Hz, employing identical increment. The local Θ-estimate of the attraction basin size is shown Figure 11, together with the largest tolerable error $R^{\mathfrak{u}}_{m_0}$. We investigate two directions: the Marmousi one of Figure 4(d) and the large reflector of Figure 4(c). We picture curves and histograms, that correspond to sequential or group of frequencies respectively. Numerical values for the attraction basis are prescribed Table 2.



(a) Evolution of $\Delta^{\mathfrak{u}}_{m_0}$ using local Θ-estimate. The large blue bars ( ) and the red dotted line ( ) employ the direction $\mathfrak{u}_m$. The smaller orange bars ( ) and the black dotted line ( ) employ direction $\mathfrak{u}_3$.

(b) Evolution of $R^{\mathfrak{u}}_{m_0}$, the large blue bars ( ) and the red dotted line ( ) employ the direction $\mathfrak{u}_m$ The smaller orange bars ( ) and the black dotted line ( ) employ direction $\mathfrak{u}_3$.

Figure 11: Evolution of $\Delta^{\mathfrak{u}}_{m_0}$ and $R^{\mathfrak{u}}_{m_0}$ with sequential or group of frequency. We use m_0 as a smooth velocity background of Figure 3. The histograms represent the estimations for groups of ten frequencies and cover the respective interval of frequency. The lines represent the estimations using sequential frequency. The different directions $\mathfrak{u}$ are either extracted from the Marmousi model ($\mathfrak{u}_m$, Figure 4(d)) or from one large reflector ($\mathfrak{u}_3$, Figure 4(c)).

The comparison provided in Figures 11(a), 11(b) and Table 2 highlight that

- the size of the attraction basin when using group of frequencies is dictated by the largest frequency in the group. Therefore, it is better to use the lowest frequency sequentially in order to have the largest size possible. For instance, using the single frequency 0.1 Hz provides a size of interval one order of magnitude larger than when including frequency up to 1 Hz.
- At a fixed frequency, taking into account lower frequencies does not affect the size of the interval.

This preference on sequential over multiple frequency is actually advocated in [15], motivated, as the frequency progression in [16], by the intuition that it reduces the cycle-skipping effect. Here

sequential frequency	$\Delta^{\mathfrak{u}_m}_{m_0}/\|m_0\|$	**frequency group**	$\Delta^{\mathfrak{u}_m}_{m_0}/\|m_0\|$
0.1 Hz	1.25	$\{0.1, 0.2, \ldots, 1\}$ Hz	0.10
1.0 Hz	8.2×10^{-2}	$\{1.1, 1.2, \ldots, 2\}$ Hz	5.3×10^{-2}
1.1 Hz	7.5×10^{-2}	$\{2.1, 2.2, \ldots, 3\}$ Hz	3.2×10^{-2}
2.0 Hz	4.5×10^{-2}	$\{3.1, 3.2, \ldots, 4\}$ Hz	2.5×10^{-2}
2.1 Hz	4.2×10^{-2}	$\{4.1, 4.2, \ldots, 5\}$ Hz	2.0×10^{-2}

Table 2: Evolution of $\Delta^{\mathfrak{u}}_{m_0}$, computed with local Θ-estimate, with frequency depending on the use of sequential or group of frequencies. Here $\mathfrak{u}$ is chosen to be the Marmousi direction, $\mathfrak{u} = \mathfrak{u}_m$, according to the Figure 4(d).

we justify this choice by a quantitative estimation of the size of the attraction basin. Moreover, we show that once the lowest frequencies are taken individually, incorporating them in higher frequency content does not modify the size of the attraction basin.

Regarding the largest tolerable error on data, Figure 11(b) shows that adding frequency increases the robustness in the direction $\mathfrak{u}_m$, except for low frequency. Then, for the direction of the large contrasting object, $\mathfrak{u}_3$, we do not observe any improvement, and the allowed distance remains relatively stable, as it was observed in Figure 5.

4.5 Selection of complex frequencies

The use of complex frequencies (also referred to as the "Laplace-Fourier" domain) has shown advantageous behavior when used during reconstruction algorithm, as mentioned in the introduction. In this case, the angular frequency ω in (2) writes as

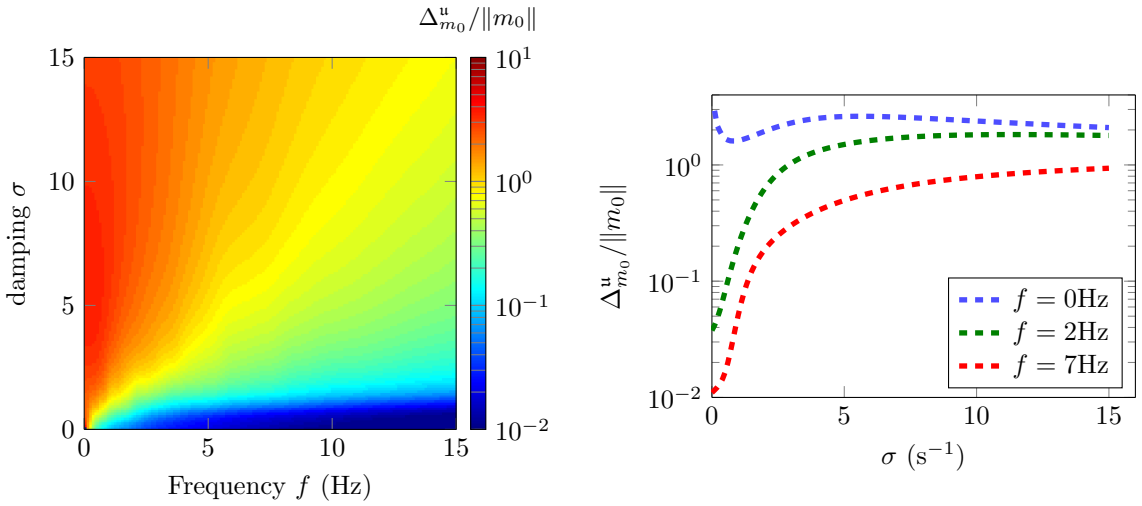
$$-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2, \qquad \mathrm{i}\omega = \sigma + 2\mathrm{i}\pi f, \tag{49}$$

where f is the 'Fourier' component of the frequency (in Hz) and σ a damping factor, representative of the Laplace component. Here, we are interested in deciding how to select the progression of those coefficients to obtain the best convergence properties. Applications have been carried out with a progression that appears mostly intuitive in literature. In the case of zero-Fourier frequency ($f = 0$), it is proposed from low to high damping coefficients, [53]. However, for the complex frequency progression ($f \neq 0$ and $\sigma \neq 0$), the evolution of the damping parameter at fixed f is proposed to be from high to low damping, see [56, 45].

We compute the estimation of $\Delta^{\mathfrak{u}}_{m_0}$ (using local Θ-estimate) and $R^{\mathfrak{u}}_{m_0}$ with complex frequencies to see how it affects the basin of attraction and robustness to noise. We consider m_0 the smooth background velocity and the similar setup as employed in Subsection 4.1. The estimates are shown in Figures 12, 13 and 14, where we compare the behaviours on a logarithmic scale for the two-dimensional complex frequency plane. Here, both f and σ varies from 0 to 15 (excluding the case where both f and σ are zero).

It is clear that the incorporation of a damping coefficient in the frequency is beneficial for the convergence, because it increases the size of the attraction basin of several orders of magnitude. Then, the process of frequency selection, for convergence, is the following:

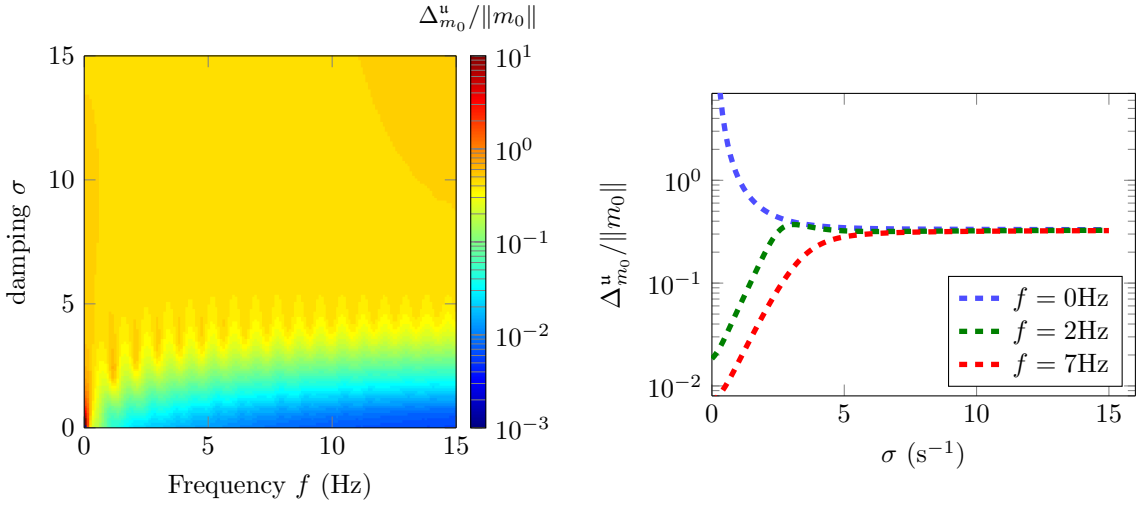
- for a fixed f, the interval size increases with increasing damping, with the exception of the case $f = 0$, cf. Figures 12(b) and 13(b).
- At fixed damping σ, the interval size decreases with increasing f.
- Following a large to small interval size gives,
  1. initial iterations should use $f = 0$ (largest interval), and a damping coefficient progression from low to high (see Figures 12(b) and 13(b)).
  2. Fix $f \neq 0$ as small as possible, the damping coefficient should evolve from high to low (see Figures 12(a) and 13(a)); then f increases progressively.
  3. Finally, one can take $\sigma = 0$ and f evolves from low to high.



(a) Estimates for the complex frequency plane.

(b) Estimates using fixed f.

Figure 12: Map of the local Θ-estimate of $\Delta^{\mathfrak{u}}_{m_0}$ with complex frequencies using m_0 as the smooth background velocity of Figure 3 and the direction $\mathfrak{u}_m$ of Figure 4(d). The angular frequency is defined by $-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2$.



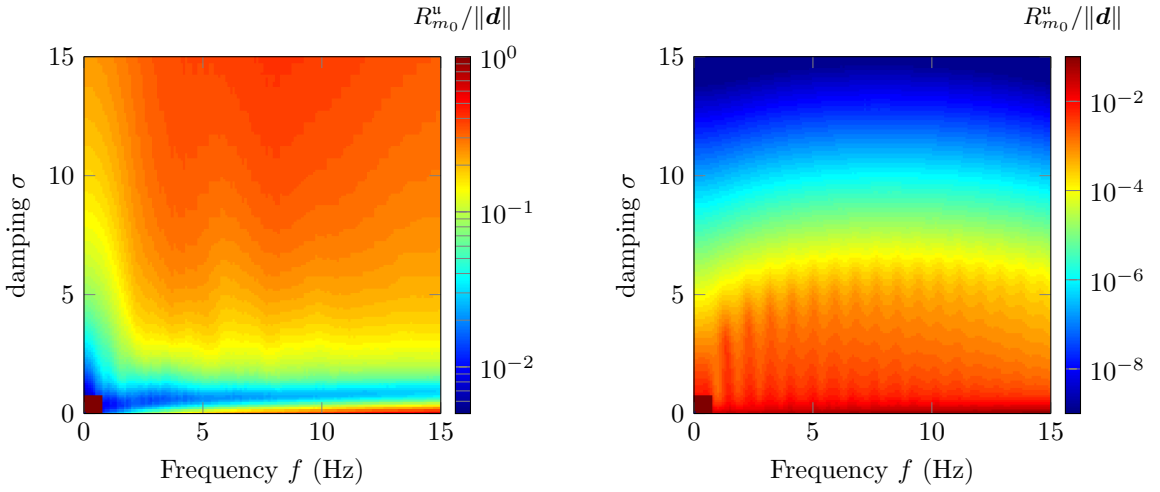
(a) Estimates for the complex frequency plane.

(b) Estimates using fixed f.

Figure 13: Map of the local Θ-estimate of $\Delta^{\mathfrak{u}}_{m_0}$ with complex frequencies using m_0 as the smooth background velocity of Figure 3 and the direction $\mathfrak{u}_3$ of Figure 4(c). The angular frequency is defined by $-\omega^2 = (\sigma+2\mathrm{i}\pi f)^2$.

Regarding the largest tolerable data error $R^{\mathfrak{u}}_{m_0}$, we see that

- at fixed σ, increasing f allows an increase in the distance between the data and the attainable set.
- The pattern depends strongly on the type of the search direction geometry, as it has already been observed in the previous experiment (and contrary to $\Delta^{\mathfrak{u}}_{m_0}$). The distance is allowed to increase with σ for the direction $\mathfrak{u}_m$, but decreases in the direction of the strong reflection, $\mathfrak{u}_3$. Also in the direction of the Marmousi reflectors, Figure 14(a), we observe a band where the allowed distance decreases, around $\sigma = 1$. On the other hand, for both directions, the case where $\sigma = 0$ appears the most robust (i.e. where the allowed distance between the data and the attainable set is maximal).



(a) Estimates using u_m.

(b) Estimates using u_3.

Figure 14: Evolution of the tolerable error level $R^{\mathrm{u}}_{m_0}$ with complex frequencies using m_0 as the smooth background velocity of Figure 3. The angular frequency is defined by $-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2$.

– Regarding the magnitude, we clearly see that the high contrast object has globally much less allowance, except when the damping is close to zero.

It is important to notice that the use of complex frequency, especially when $f = 0$, improves the convergence by increasing notably the size of the attraction basin. Therefore, it provides a fundamental benefit for the reconstruction, in particular when no information is initially known. Note that it basically acts as a very low frequency (note that the magnitude of the estimates for complex frequencies eventually matches the very low frequency ones; we can see a continuation in the magnitudes). On the other hand, such frequencies are more sensitive to noise, and require accuracy in the data. We refer to [53, 56, 45, 31] for illustrations of reconstruction.

Remark 4 (Stability and frequency progression)**.** *We have shown that the reconstruction procedure requires low frequencies to ensure a larger size for the basin of attraction, hence allowing the better chance of convergence despite the possible lack of prior information. However, higher frequencies are still required for reasons of stability, and justify the traditional progression in frequency along with the minimization iterations. Stability of inverse problem has been studied in particular with the Calderón problem, where the data are the Dirichlet-to-Neumann map. Assumptions are commonly introduced to obtain a* conditional *Lipschitz-type stability results, see, e.g., [2, 40, 3, 12] and the references therein. For instance, the piecewise constant model representation introduced Remark 1 is a necessary ingredient, e.g., [3, 11]. Lipschitz-type stability is obtained when, for two models m_0 and $m_\dagger$, it holds that*

$$\|m_0 - m_\dagger\| \leq \mathcal{C}\|\mathcal{F}(m_0) - \mathcal{F}(m_\dagger)\|, \tag{50}$$

where $\mathcal{C}$ represents the stability constant. The iterative procedure for the reconstruction minimizes the difference between the forward map and the stability result ensures the accuracy of the reconstruction (depending on the values of $\mathcal{C}$).

In [11], numerical estimates of the stability constant are obtained in the context of the geophysical inverse problem. It illustrates, for example, the influence of the number of coefficients n (in the model representation (7)) for stability. When the stability constant $\mathcal{C}$ is small, it indicates that the correspondence in the data guarantees the accuracy of the reconstruction. Here, we estimate the stability constant (following the methodology of [11]) to see its dependency with complex frequency, in order to compare with the convergence estimates of Figures 12 and 13. We use m_0 the smooth background of Figure 3 and select $m_\dagger$ to be the Marmousi model of Figure 8(a). The resulting stability constant estimates are pictured Figure 15. We observe that

– *the stability constant estimates decrease (i.e. the expected accuracy of the reconstruction is improved) with increasing frequency f.*

– *Incorporating damping with frequency deteriorates the stability constant (the numerical estimates increases). More precisely, low frequencies and damping (f and σ respectively) give the worst result in terms of stability.*

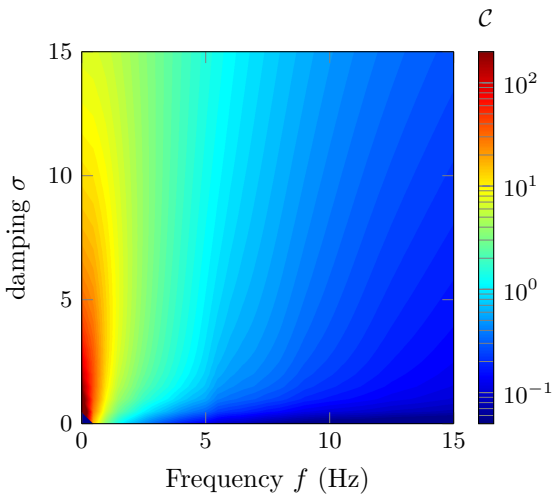


Figure 15: Numerical estimates for the stability constant $\mathcal{C}$ in (50) with complex frequencies using m_0 as the smooth background velocity of Figure 3 and $m_\dagger$ as the Marmousi model of Figure 8(a). The angular frequency is defined by $-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2$.

Therefore, the estimates for convergence (Figures 12 and 13) and stability (Figure 15) behave oppositely. The convergence improves at low frequency and with damping meanwhile the stability deteriorates. Higher frequency reduces the convergence radius but ameliorates the stability (by reducing the wavelength, such frequencies allow to capture more details). Both aspects of the problem have to be taken into account during the reconstruction; it designs the well-known multi-frequency algorithm for the computation, starting from low frequency (to foster the convergence) towards high frequency (to foster stability) (see, e.g., Algorithm 1).

5 Optimizability: MBTT versus classical FWI

In this section, we apply the optimizability analysis tools of Section 3 to compare the MBTT formulation of FWI (18) (where the unknown model is parametrized by a smooth background $\mathfrak{p}$ and a data space reflectivity $\mathfrak{s}$), with the original FWI problem (13) (where the unknown model is the squared slowness m). The MBTT approach was originally designed to overcome the phase shift problem which hampers classical FWI, so the optimizability study of this section is expected to quantify the gain - if any - of the MBTT formulation over the classical FWI in this respect.

5.1 Choice of the model

For a fair comparison, the two approaches have to be applied to *the same model*, so we construct a nominal model $\mathfrak{m}_0$ whose MBTT decomposition, $\mathfrak{p}_0$, $\mathfrak{s}_0$, is known exactly, i.e. which satisfies $\mathfrak{m}_0 = m(\mathfrak{p}_0, \mathfrak{s}_0)$ according to (15). We have decided to construct a model $\mathfrak{m}_0$ inspired by the Marmousi model m_m of Figure 8(a). This is achieved by defining $\mathfrak{p}_0$ and $\mathfrak{s}_0$ by:

$$\mathfrak{p}_0 = m_0 \quad , \quad \forall \omega \ : \ \mathfrak{s}_0(\omega) = \mathcal{F}_\omega(m_m) - \mathcal{F}_\omega(m_0), \tag{51}$$

where m_m is the Marmousi model of Figure 8(a) and m_0 is the smooth background of Figure 3. Then, $\mathfrak{m}_0$ is defined by:

$$\mathfrak{m}_0 \stackrel{\text{def}}{=} \mathfrak{p}_0 + \sum_\omega \underbrace{\mathcal{W}(\omega) D\mathcal{F}^*_\omega(\mathfrak{p}_0)\mathfrak{s}_0(\omega)}_{\mathfrak{r}_0(\omega)} = \mathfrak{p}_0 + \underbrace{\mathcal{W} D\mathcal{F}^*(\mathfrak{p}_0)\mathfrak{s}_0}_{\mathfrak{r}_0} = m(\mathfrak{p}_0, \mathfrak{s}_0), \tag{52}$$

where $\mathfrak{r}_0(\omega)$ (respectively $\mathfrak{r}_0$) is the depth reflectivity associated to frequency ω[1] (respectively to the sum of all frequencies). Definition (52) guaranties that:

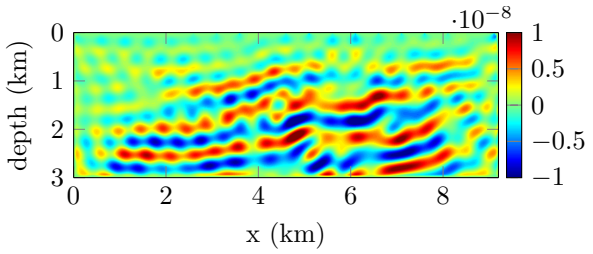
$$\mathcal{F}(\mathfrak{m}_0) = \mathsf{F}(\mathfrak{p}_0, \mathfrak{s}_0). \tag{53}$$

We choose the weight $\mathcal{W}$ proportional to the square root of depth, as proposed in Section 2, with a proportionality coefficient such that the *model reflectivity level* β *at frequency* ω defined by

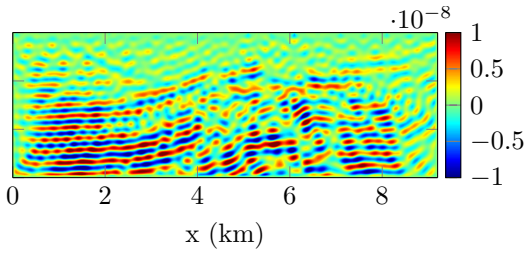
$$\beta(\omega) \stackrel{\text{def}}{=} \|\mathfrak{r}_0(\omega)\|/\|\mathfrak{p}_0\| \tag{54}$$

takes the value 1% at all frequencies. When β is small, the forward modeling part of the MBTT approach becomes close to the Born approximation.

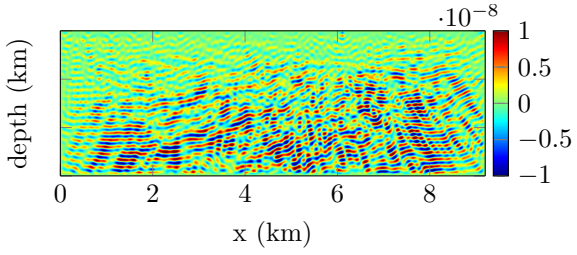
In Figure 16, we illustrate the resulting models $\mathfrak{r}_0(\omega)$ for three frequencies: 2, 4 and 7 Hz. We also show the model $\mathfrak{r}_0$ where the frequency sum contains frequencies between 0.5 to 15 Hz, with 0.5 Hz increment. We observe that the reflectivity, defined from the difference between observations and simulations using a smooth background, provides structures of size consistent with the selected frequency. For the global model, shown Figure 16(d), we see the contributions of all wavelengths, and we can distinguish some structures of the Marmousi medium given Figure 8(a).



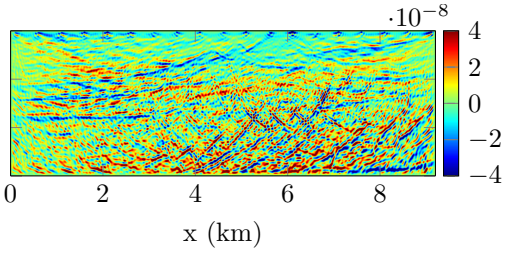
(a) Model $\mathfrak{r}_0$ at 2 Hz frequency.



(b) Model $\mathfrak{r}_0$ at 4 Hz frequency.



(c) Model $\mathfrak{r}_0$ at 7 Hz frequency.



(d) Model $\mathfrak{r}_0$ using frequencies from 0.5 to 15 Hz with 0.5 Hz increment.

Figure 16: Reflectivity models $\mathfrak{r}_0$ obtained from the MBTT representation defined by (52). The model $\mathfrak{s}$ is defined from (51) as the difference between the data obtained from the Marmousi model Figure 8(a) and the smooth background Figure 3. The figures correspond with squared slowness and the values are given in $(\mathrm{m\,s^{-1}})^{-2}$.

For simplicity, in the following, we restrict ourselves by studying only *single frequency* nominal models, which means that we only work with models resulting from a single, fixed ω:

$$\mathfrak{m}_0(\omega) = \mathfrak{p}_0 + \mathcal{W}(\omega) D\mathcal{F}^*_\omega(\mathfrak{p}_0)\mathfrak{s}_0(\omega) = \mathfrak{p}_0 + \mathfrak{r}_0(\omega). \tag{55}$$

[1]Note that with $\mathfrak{s}_0$ given by (51), the resulting $\mathfrak{r}_0(\omega)$ is the gradient of the ($L2$ norm) cost function at frequency ω using the model $\mathfrak{p}_0$, see Appendix A.

It allows us to study the behaviour of both approaches (global velocity model reconstruction or MBTT decomposition) with individual frequency, preventing artifacts that may appear due to the frequency summation in the reflectivity.

5.2 Choice of perturbation directions

Background perturbation We consider a background perturbation in the direction of the one-dimensional ramp $\mathfrak{u}$ of Figure 17, and illustrate the effect of such a perturbation applied either to the global model m, as in the case of FWI, or to the background unknown $\mathfrak{p}$ in the case of MBTT.

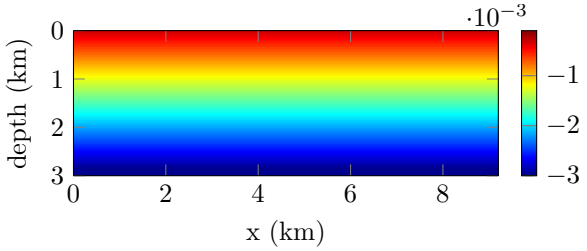


Figure 17: Perturbation $\mathfrak{u}$ used for the background model $\mathfrak{p}$. The amplitude is determined such that $\|\mathfrak{u}\| = 1$ and the values are given in $(\mathrm{m\,s^{-1}})^{-2} = \mathrm{s^2\,m^{-2}}$.

Figure 18 shows the corresponding unperturbed and perturbed synthetic data for a single, centrally located source at frequency 4 Hz. It corresponds to the solution of the Helmholtz equation (2) recorded at the receivers. Note that, from (10) the direct arrivals are removed from the forward operator. We see that the modification of the background part $\mathfrak{p}$ preserves the phase of the original signal, and only the amplitudes are modified. On the other hand, when the perturbation is applied to the global model m, both phase and amplitude of the signal are modified.

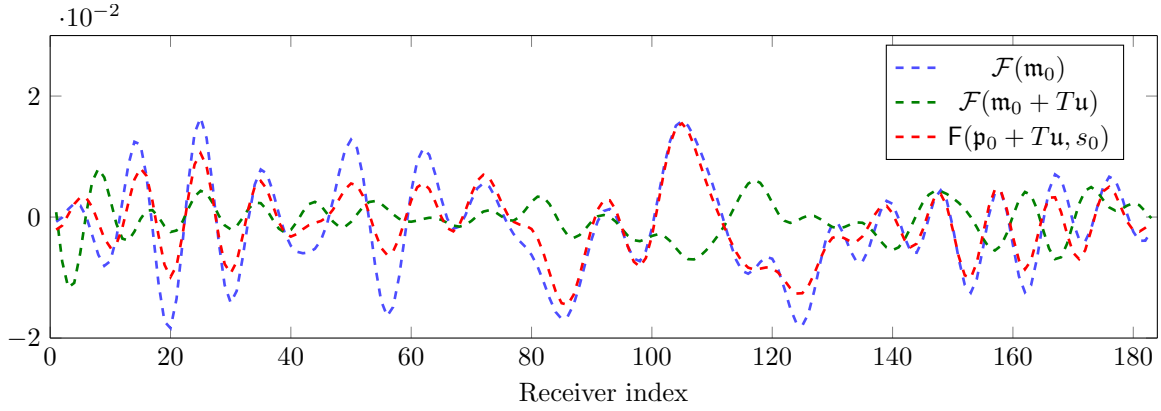


Figure 18: Comparison of the synthetic data associated with a centrally located shot at 4 Hz using a model perturbed by the direction $\mathfrak{u}$ of Figure 17 applied onto the global model m or on the part $\mathfrak{p}$ using the MBTT model decomposition. The step for the perturbation is $T = 5 \times 10^{-5}$.

Reflectivity perturbation The FWI objective function is known to be nearly quadratical with respect to reflectivity, i.e. to the high spatial frequency part of m, and the same property holds by construction for the dependance of the MBTT objective function with respect to the data space reflectivity $\mathfrak{s}$. Hence one expects large basins of attraction with respect to $\mathfrak{s}$ in the MBTT formulation. We choose for $\mathfrak{u}_\mathfrak{s}$ a random vector of the data space.

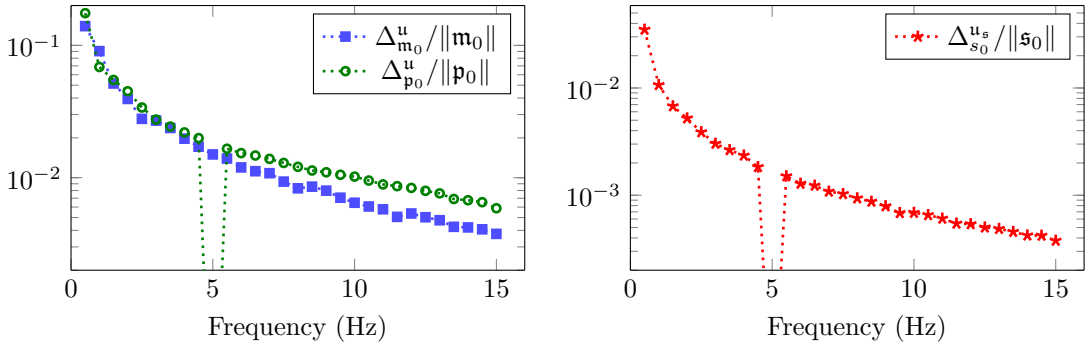
5.3 Comparison of Local Θ-estimates

The formulas for local Θ-estimates have already been derived in Section 3.3 for classical FWI. Application of these formulas to $\mathsf{F}(\mathfrak{p},\mathfrak{s})$ instead of $\mathcal{F}(m)$ gives immediately the local Θ-estimates of the sizes $\Delta^{\mathfrak{u}}_{\mathfrak{p}_0}$ and $\Delta^{\mathfrak{u}_\mathfrak{s}}_{\mathfrak{s}_0}$ of attraction basins for $\mathfrak{p}$ and $\mathfrak{s}$ at $\mathfrak{p}_0$, $\mathfrak{s}_0$ in directions $\mathfrak{u}$, $\mathfrak{u}_\mathfrak{s}$ for the MBTT formulation of FWI:

$$\begin{cases} \Delta^{\mathfrak{u}}_{\mathfrak{p}_0} = \dfrac{\pi}{4}\dfrac{\|D_\mathfrak{p}\mathsf{F}(\mathfrak{p}_0,\mathfrak{s}_0)(\mathfrak{u})\|}{\|D^2_\mathfrak{p}\mathsf{F}(\mathfrak{p}_0,\mathfrak{s}_0)(\mathfrak{u},\mathfrak{u})\|}, \\ \Delta^{\mathfrak{u}_\mathfrak{s}}_{\mathfrak{s}_0} = \dfrac{\pi}{4}\dfrac{\|D_\mathfrak{s}\mathsf{F}(\mathfrak{p}_0,\mathfrak{s}_0)(\mathfrak{u}_\mathfrak{s})\|}{\|D^2_\mathfrak{s}\mathsf{F}(\mathfrak{p}_0,\mathfrak{s}_0)(\mathfrak{u}_\mathfrak{s},\mathfrak{u}_\mathfrak{s})\|}. \end{cases} \tag{56}$$

We have omitted the frequency index for clarity in the notation. Note that this is not misleading because we have chosen to work with single frequency content here, i.e., the reflectivity $\mathfrak{r}$ only contains one selected frequency.

Figure 19 shows the evolution of the local Θ-estimates with frequency. The size of the corresponding attraction basins decreases with frequency, when the perturbation is applied on m, $\mathfrak{p}$ and $\mathfrak{s}$. The left part of the figure shows a slightly larger attraction basin in the direction of the background perturbation $\mathfrak{u}$ when it is applied to the propagator part $\mathfrak{p}_0$ of the MBTT parameterization rather than when applied directly to $\mathfrak{m}_0$. But its not up to the expectations raised by the claim that the MBTT parameterization allows to overcome the phase shift problem [20, 21]. Regarding $\mathfrak{s}$, the estimated size appears surprisingly small compared to the large attraction basin expected, see Section 5.2. But one has to remember that these are *local* Θ-estimate, which can be very pessimistic, as explained in Section 3.3, and postpone more definitive comments to the end of Subsection 5.4, where exact Θ-estimates are calculated.



(a) Perturbation of the background model $\mathfrak{p}$.

(b) Perturbation of the reflectivity $\mathfrak{s}$.

Figure 19: Evolution with frequency of the local Θ-estimates of the size of the attraction basins in the context of FWI or MBTT model space parametrization, see (42) and (56). Here $\mathfrak{p}_0$ is the smooth velocity background of Figure 3, the direction $\mathfrak{u}$ for $\mathfrak{p}$ is given Figure 17, and the direction $\mathfrak{u}_\mathfrak{s}$ for $\mathfrak{s}$ is a random vector. In the MBTT representation, the reflectivity uses only the selected frequency.

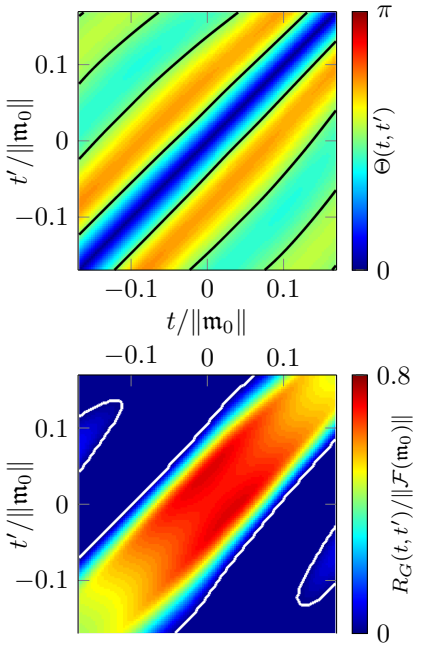
5.4 Comparison of exact Θ- and R_G-estimates

We apply the method described in Section 3.4 for the case of classical FWI, which translates immediately to the case of MBTT by replacing the FWI forward map $m \rightsquigarrow \mathcal{F}(m)$ by the MBTT forward map $\mathfrak{p},\mathfrak{s} \rightsquigarrow \mathsf{F}(\mathfrak{p},\mathfrak{s})$. This leads to the computation of *deflection* and *global radius of curvature maps* between the following points:

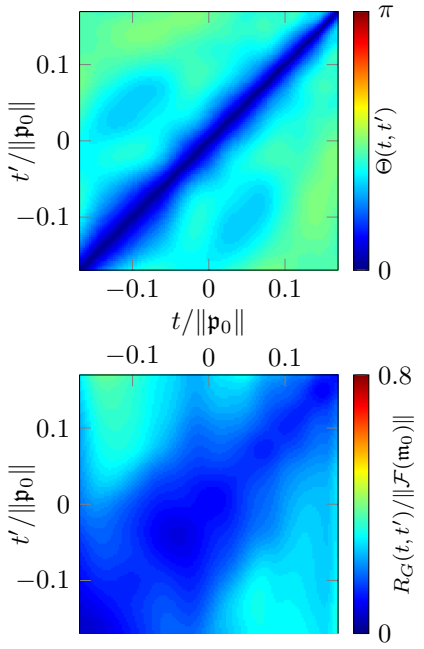
FWI	(attraction basin for $\mathfrak{m}$)	:	$\mathcal{F}(\mathfrak{m}_0+t\mathfrak{u})$	and	$\mathcal{F}(\mathfrak{m}_0+t'\mathfrak{u})$;
MBTT	(attraction basin for $\mathfrak{p}$)	:	$\mathsf{F}(\mathfrak{p}_0+t\mathfrak{u},\mathfrak{s}_0)$	and	$\mathsf{F}(\mathfrak{p}_0+t'\mathfrak{u},\mathfrak{s}_0)$;
MBTT	(attraction basin for $\mathfrak{s}$)	:	$\mathsf{F}(\mathfrak{p}_0,\mathfrak{s}_0+t\mathfrak{u}_\mathfrak{s})$	and	$\mathsf{F}(\mathfrak{p}_0,\mathfrak{s}_0+t'\mathfrak{u}_\mathfrak{s})$;

for all $-\Delta \le t,t' \le \Delta$.

We first compute the deflection and global radius of curvature maps for $\mathfrak{m}$ and $\mathfrak{p}$, using values of t and t' in an interval $[-\Delta, \Delta]$ which is chosen to represent in each case about $\pm 20\%$ of the norm of $\mathfrak{m}_0$ or $\mathfrak{p}_0$ defined in (52). Figures 20 and 21 show these maps at two selected frequencies: 4 and 7 Hz, and Table 3 summarizes the extracted exact estimates of $\Delta^{\mathfrak{u}}_{\mathfrak{m}_0}$, $\Delta^{\mathfrak{u}}_{\mathfrak{p}_0}$, together with the local estimates extracted from Figure 19(a).



(a) FWI: perturbation of the global model $\mathfrak{m}$.



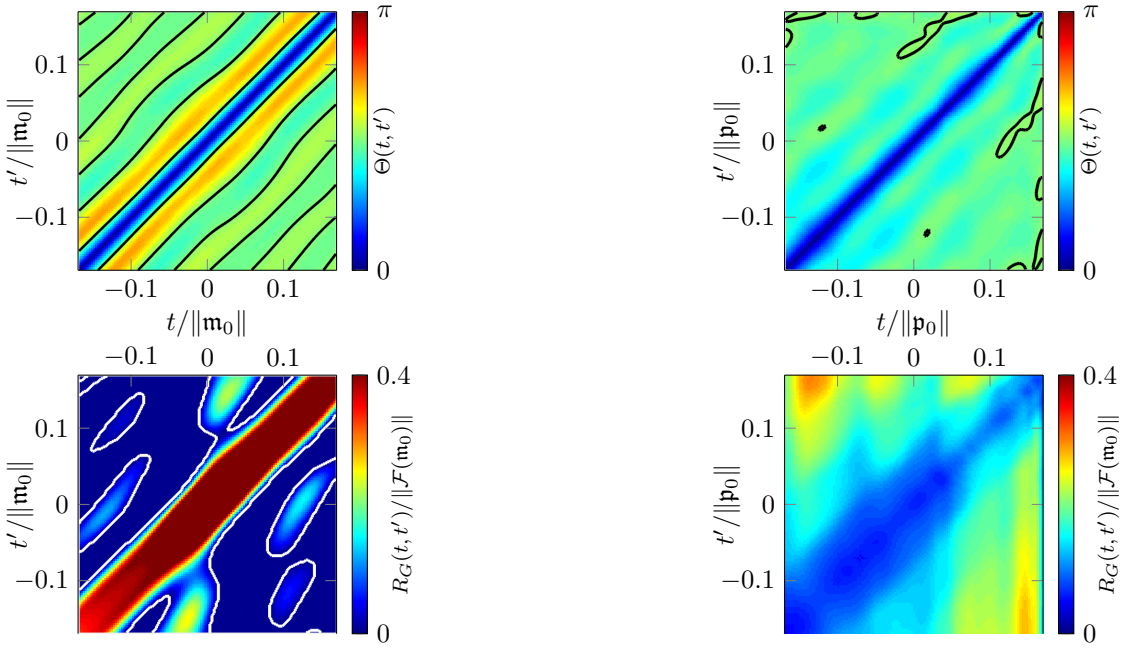
(b) MBTT: perturbation of the background $\mathfrak{p}$.

Figure 20: Maps of the deflection (30) and global radius (26) between two perturbed velocity or background models at frequency 4 Hz. The perturbation direction is the ramp of Figure 17, it is either applied to the global model $\mathfrak{m}$ or to the background parameter $\mathfrak{p}$. The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

	model m		model $\mathfrak{p}$		model $\mathfrak{s}$	
	$\frac{\Delta^{\mathfrak{u}}_{\mathfrak{m}_0}}{\|\mathfrak{m}_0\|}$	$\frac{R^{\mathfrak{u}}_{G,\mathfrak{m}_0}}{\|\mathcal{F}(\mathfrak{m}_0)\|}$	$\frac{\Delta^{\mathfrak{u}}_{\mathfrak{p}_0}}{\|\mathfrak{p}_0\|}$	$\frac{R^{\mathfrak{u}}_{G,\mathfrak{p}_0}}{\|\mathcal{F}(\mathfrak{m}_0)\|}$	$\frac{\Delta^{\mathfrak{u}}_{\mathfrak{s}_0}}{\|\mathfrak{s}_0\|}$	$\frac{R^{\mathfrak{u}}_{G,\mathfrak{s}_0}}{\|\mathcal{F}(\mathfrak{m}_0)\|}$
4 Hz						
Local Θ-estimates	0.02	1.6	0.022	0.8	2×10^{-3}	6.5
Exact Θ-estimates	0.02	0.6	0.2	0.05	54	6.5
Exact R_G-estimates	0.05	0.0	0.23	0	60	0
7 Hz						
Local Θ-estimates	0.01	1.6	0.014	0.7	1×10^{-3}	4.9
Exact Θ-estimates	0.01	0.6	0.11	0.06	23	4.9
Exact R_G-estimates	0.025	0.0	0.20	0	>35	0

Table 3: Size Δ of attraction basins centered at $\mathfrak{m}_0$ and corresponding maximal tolerable error R_G for the different estimations, at 4 and 7 Hz. By construction, the R_G-estimates correspond to the limit case of a zero tolerable error, the other values are extracted from Figures 19, 20, 21 and 22.

– The first observation is that lower values of deflection are achieved when the background perturbation $\mathfrak{u}$ is applied to $\mathfrak{p}$ (MBTT) rather than to $\mathfrak{m}$ (FWI): at 4 Hz, Figures 20(b), it never



(a) FWI: perturbation of the global model m.

(b) MBTT: perturbation of the background $\mathfrak{p}$.

Figure 21: Maps of the deflection (30) and global radius (26) between two perturbed velocity or background models at frequency 7 Hz. The perturbation direction is the ramp of Figure 17, it is either applied to the global model $\mathfrak{m}$ or to the background parameter $\mathfrak{p}$. The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

reaches $\pi/2$, and at 7 Hz, Figure 21(b), only a few portions attain this value. On the contrary, for FWI, Figures 20(a) and 21(a), the deflection rapidly reaches $\pi/2$ at both frequencies. This indicates that the MBTT formulation produces larger Θ-attraction basins than the standard FWI formulation, roughly by a factor ten (Table 3). Notice that the size of the Θ-attraction basin is divided by two for both FWI and MBTT when the frequency increases from 4 to 7 Hz.

– The second observation concerns the *strict positivity* of the global radius of curvature R_G, bottom of Figures 20 and 21, which determines the R_G-attraction basin characterized by a zero tolerable error (see Section 3.4). For the MBTT formulation, Figures 20(b) and 21(b), R_G remains strictly positive all over the map, which shows that the R_G-basin is larger than the investigated interval. Its size is of approximately 20% at both 4 and 7 Hz (Table 3). On the contrary, for the usual FWI formulation, Figures 20(a) and 21(a), R_G decreases very rapidly to zero when one moves away of the diagonal, producing smaller R_G-attraction basins, with size of 5% at 4 Hz and 2.5% at 7 Hz, smaller by a factor four to eight to the corresponding MBTT attraction basins.

– Concerning the *magnitude* of R_G, whose minimum over the attraction basin gives the tolerable error level, one sees that it takes *larger values* for FWI near the main diagonal, i.e. for small attraction basins, than for MBTT over the whole map, i.e. for large attraction basins, which is confirmed by the values of R_G in Table 3, which give the tolerable error level associated with the Θ-attraction basins (this level is zero by definition for the R_G attraction basins).

To summarize, MBTT extends significantly the size of attraction basins with respect to background perturbations, at the price of a reduction in the admissible error level. This explains the success of MBTT's alternate minimization algorithm, as reported in [22].

We compare now the above exact Θ-estimates with the local Θ-estimates of Subsection 5.3.

– For the FWI approach, Figures 19(a), 20(a) and 21(a) and Table 3, it shows that both *local* and *exact* Θ-estimates of $\Delta^{\mathfrak{u}}_{\mathfrak{m}}$ are of the same size. In sight of the upper bound estimate (32) on which the local Θ-estimate is based, one can think that the FWI formulation corresponds to the worst situation, where the image of a segment in the background space is a curve close to an arc of circle in the data space.

– The situation is completely different for the MBTT formulation: Figures 19(a), 20(b) and 21(b) and Table 3, we see that the exact Θ-estimate $\Delta^{\mathfrak{u}}_{\mathfrak{p}}$ is about ten times larger than its *local* Θ-estimate.

We can also determine the exact attraction basins in the MBTT formulation for the data space reflectivity $\mathfrak{s}$ at $\mathfrak{s}_0$ in the direction $\mathfrak{u}_\mathfrak{s}$, which we expect to be large because the forward map F is nearly linear with respect to $\mathfrak{s}$. Figure 22 shows the corresponding deflection and global radius of curvatures maps for values of t and t' in an interval $[-\Delta, \Delta]$ which is chosen to represent in each case about ± 35 times the norm of $\mathfrak{s}_0$ defined in (52). As expected, the exact Θ-attraction basin is large (23 to 54 times the norm of $\mathfrak{s}_0$ depending on frequency), and is 10^5 times larger than its local estimate, which, together with the previous results on the estimation of $\Delta^{\mathfrak{u}}_{\mathfrak{p}}$, confirms the necessity of exact estimates for accuracy.

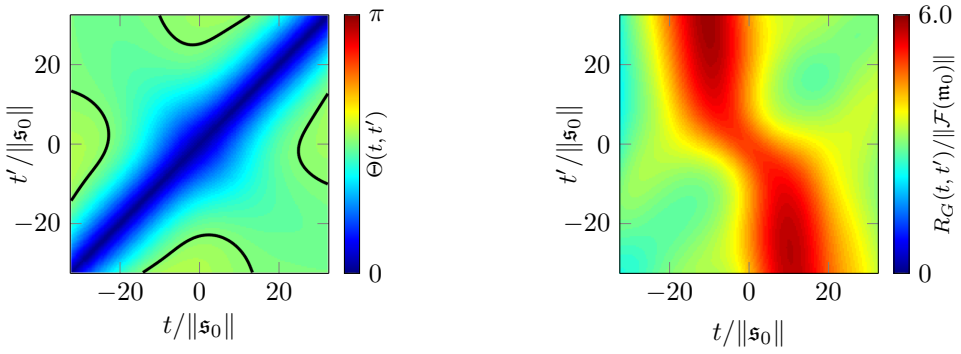
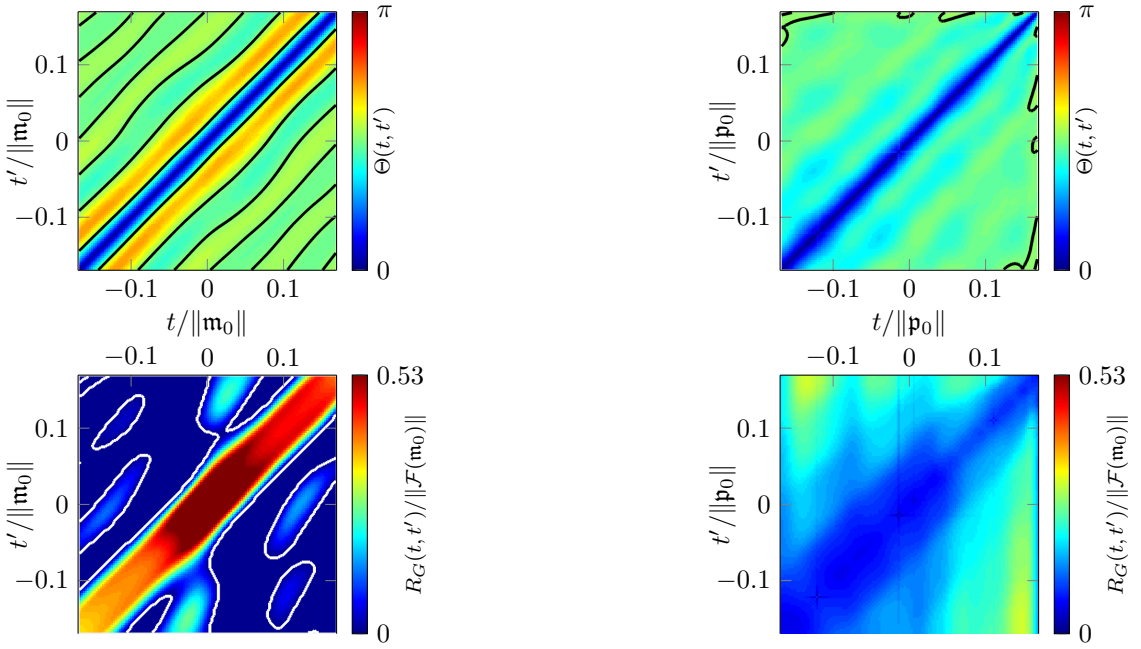


Figure 22: Maps of the deflection (30) (left) and global radius (26) (right) at frequency 7 Hz between two perturbed data-space reflectivities, for the perturbation direction $\mathfrak{u}_\mathfrak{s}$ chosen in section 5.2 The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

5.5 Influence of reflectivity level

The previous comparisons at 4 Hz and 7 Hz have been performed on the nominal model which has a reflectivity level β of 1% (see (54)). We repeat the same computations, at 7 Hz, changing only the reflectivity of the model, which is fixed to a low level, 0.2% in Figure 23 (close to the Born approximation) and to a stronger value, 5% (close to Marmousi reflectivity) in Figure 24. Table 4 summarizes the sizes of and tolerable error for the corresponding attraction basins.

Increasing the reflectivity increases the relative importance of multiple reflections - whose phase are *not* controlled by the migration/demigration included in the MBTT approach - with respect to primary reflections, whose phase *are* controlled. It is then expected that the performance of MBTT, when applied to a full wave model as it is the case here, will deteriorate when the reflectivity level increases. This is confirmed by Table 4: the MBTT attraction basin for a reflectivity level of 5 % is three times smaller than at 1 %, which itself is two times smaller than at 0,2 %. On the other hand, the FWI attraction basin remains essentially unaffected by the reflectivity level, and the MBTT basins remain always larger, with a factor four to fifteen (with respect to the Θ-estimates, eight with the R_G-estimates).



(a) FWI: perturbation of the global model m.

(b) MBTT: perturbation of the background $\mathfrak{p}$.

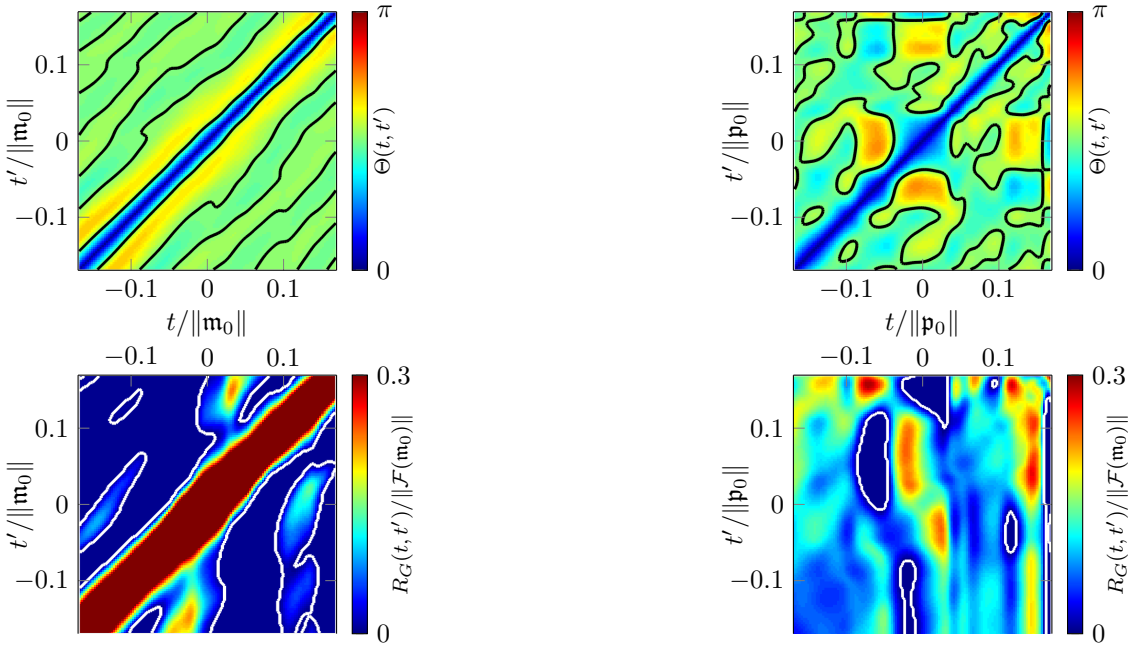
Figure 23: Maps of the deflection (30) and global radius (26) between two perturbed velocity or background models at frequency 7 Hz. Compared to Figure 21, the reflectivity β defined by (54) has been divided by five. The perturbation direction is the ramp of Figure 17, it is either applied to the global model $\mathfrak{m}$ or to the background parameter $\mathfrak{p}$. The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

7 Hz	model m		model $\mathfrak{p}$	
	$\frac{\Delta^{\mathfrak{u}}_{\mathfrak{m}_0}}{\|\mathfrak{m}_0\|}$	$\frac{R^{\mathfrak{u}}_{G,\mathfrak{m}_0}}{\|\mathcal{F}(\mathfrak{m}_0)\|}$	$\frac{\Delta^{\mathfrak{u}}_{\mathfrak{p}_0}}{\|\mathfrak{p}_0\|}$	$\frac{R^{\mathfrak{u}}_{G,\mathfrak{p}_0}}{\|\mathcal{F}(\mathfrak{m}_0)\|}$
	low reflectivity $\beta = 2 \times 10^{-3}$			
Exact Θ-estimates	0.011	0.648	0.155	0.060
Exact R_G-estimates	0.025	0	0.214	0
	medium reflectivity $\beta = 1 \times 10^{-2}$			
Exact Θ-estimates	0.011	0.645	0.107	0.060
Exact R_G-estimates	0.025	0	0.202	0
	strong reflectivity $\beta = 5 \times 10^{-2}$			
Exact Θ-estimates	0.008	0.580	0.037	0.124
Exact R_G-estimates	0.028	0	0.042	0

Table 4: Comparison of the size Δ of attraction basins centered at $\mathfrak{m}_0$ and corresponding maximal tolerable error R_G for the different levels of reflectivity at 7 Hz. By construction, the R_G-estimates correspond to the limit case of a zero tolerable error, the values are extracted from Figures 21, 23 and 24.

5.6 Influence of background smoothness

Here we illustrate an important condition which has to be satisfied for the MBTT decomposition to work. When a background perturbation is applied to $\mathfrak{p}_0$ while $\mathfrak{s}_0$ is maintained fixed, the phase shifts of the reflections are controlled, provided these reflections are generated *only* by $\mathfrak{s}_0$. Any backscattered energy generated by the updated background $\mathfrak{p}_0 + t\mathfrak{u}$ - or worse, by $\mathfrak{p}_0$ itself - will lead to reflections whose phase is not controlled by the migration-demigration process included in MBTT,



(a) FWI: perturbation of the global model m.

(b) MBTT: perturbation of the background $\mathfrak{p}$.

Figure 24: Maps of the deflection (30) and global radius (26) between two perturbed velocity or background models at frequency 7 Hz. Compared to Figure 21, the reflectivity β defined by (54) has been multiplied by five. The perturbation direction is the ramp of Figure 17, it is either applied to the global model $\mathfrak{m}$ or to the background parameter $\mathfrak{p}$. The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

thus annihilating its benefits. This appears already in Figure 21(b), where one sees that the deflection reaches $\pi/2$ (resp. the global radius of curvature equates zero) much closer to the diagonal at the top right corner than at the bottom left one. Indeed, the large positive t, t' correspond to strongly increasing backgrounds, which backscatter energy. At the opposite corner of the diagrams (bottom left), which corresponds to mildly increasing velocities, the amount of scattered energy is smaller, and the limiting values are further apart from the diagonal. We have considered for simplicity only attraction basins which are centered at the nominal model $\mathfrak{m}_0$, but on most of the deflection and curvature maps, the attraction basin could, in fact, be extended on the side of negative t (where backgrounds are less steep).

Then, if the unperturbed background itself scatters back some energy, as the one illustrated Figure 25, the situation is even worse, as illustrated in the corresponding deflection and curvature maps of Figure 26, at 4 Hz. These maps show no enlargement of the attraction basin in favor of MBTT, in fact, it is even smaller than for FWI. That is why it is fundamental to ensure the smoothness of the background in the MBTT decomposition by using an adapted parametrization, e.g., spline functions are employed in [22] to represent $\mathfrak{p}$.

5.7 Behavior of the data misfit

We illustrate in this section the behavior of the FWI and MBTT misfit functions in the model space direction $\mathfrak{u}$ at the nominal model $\mathfrak{m}_0 = m(\mathfrak{p}_0, \mathfrak{s}_0)$ for some seismic data $\boldsymbol{d}$. The cost functions are respectively $\mathcal{J}(\mathfrak{m}_0 + t\mathfrak{u})$ and $\mathsf{J}(\mathfrak{p}_0 + t\mathfrak{u}, \mathfrak{s}_0)$ for FWI and MBTT, according to (13) and (18). This will allow to verify the theoretical results of Section 3: if the minimum of the misfit function is smaller than the tolerable error $R_G^{\mathfrak{u}}$ associated to an attraction basin, the misfit should not have any local minimum over the basin. However, when the data exhibit an error larger than the tolerable error, local minima *may* appear, but they do not have to.

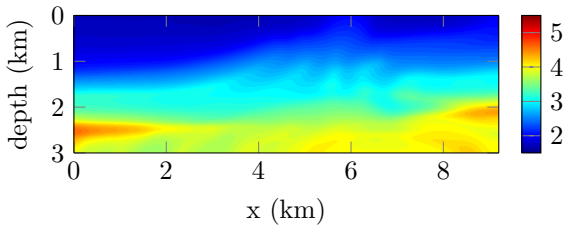
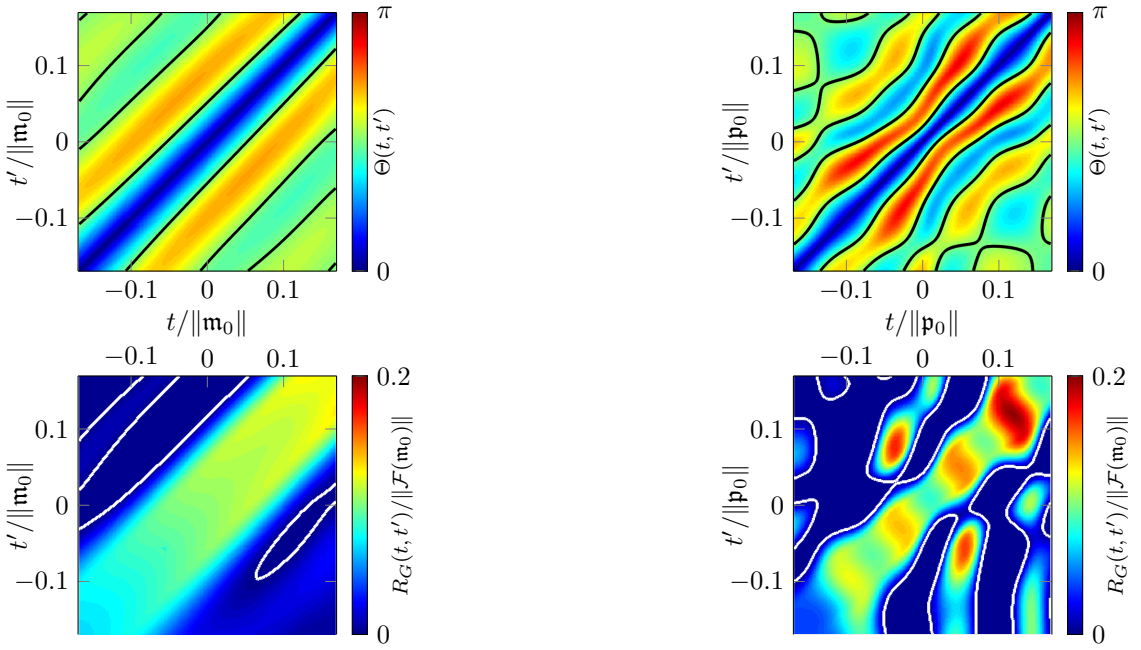


Figure 25: Acoustic smooth Marmousi model c_0 where the velocity is indicated in $\mathrm{km\,s^{-1}}$. The estimates are calculated using $\mathfrak{p}_0 = c_0^{-2}$.



(a) FWI: perturbation of the global model m.

(b) MBTT: perturbation of the background $\mathfrak{p}$.

Figure 26: Maps of the deflection (30) and global radius (26) between two models at frequency 4 Hz. The model $\mathfrak{p}_0$ is shown Figure 25 and is the only difference in the computation compared to the maps shown Figure 20. The perturbation direction is the ramp of Figure 17, and is either applied to the global model m or to the background $\mathfrak{p}$. The black lines indicate when the deflection becomes higher than $\pi/2$, the white lines indicate when the global radius becomes 0.

In order to define a seismic data $\boldsymbol{d}$ to build the misfit, we first denote by $\boldsymbol{d}_0 = \mathcal{F}(\mathfrak{m}_0) = \mathsf{F}(\mathfrak{p}_0, \mathfrak{s}_0)$ the synthetic data associated to the nominal model. Note that, due to the choices made in the construction of the nominal model in Subsection 5.1, $\boldsymbol{d}_0$ is nothing but a migrated-demigrated version of the Marmousi data deprived of the direct arrivals. We define a *scaled Marmousi data* $\boldsymbol{d}_m$ by scaling the synthetic Marmousi data $\boldsymbol{d}_M = \mathcal{F}(m_m)$ (where m_m is given Figure 8(a)) in such a way that $\|\boldsymbol{d}_m\| = \|\boldsymbol{d}_0\|$. Then, the data that enter in the FWI and MBTT misfit functionals are defined by:

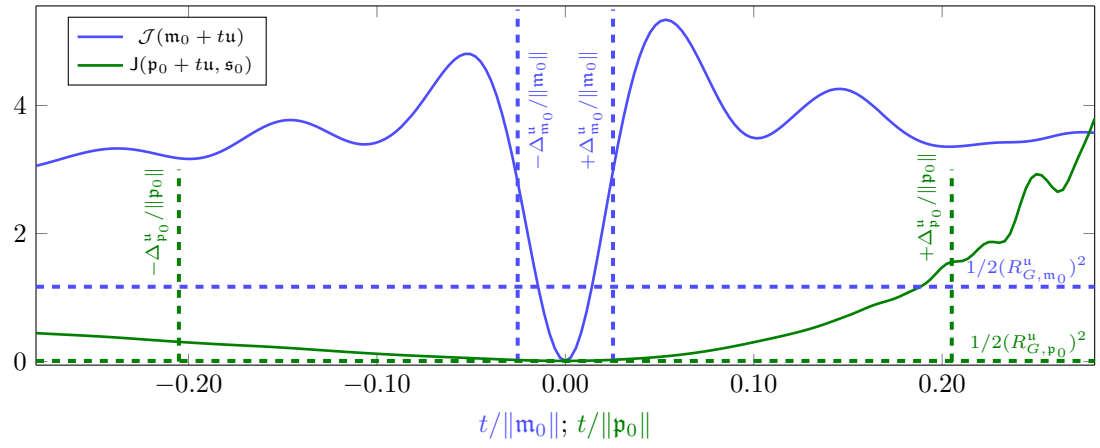
$$\boldsymbol{d} = (1-\eta)\boldsymbol{d}_0 + \eta \boldsymbol{d}_m, \qquad \eta \in [0,1]. \tag{57}$$

When η increases from 0 to 1, the distance of $\boldsymbol{d}$ to the attainable set increases from 0 to $\|\boldsymbol{d}_m - \boldsymbol{d}_0\|$, which is larger than the tolerable error associated to the attraction basin of interest (FWI or MBTT).

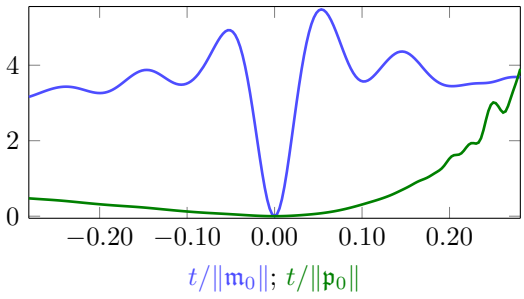
The computations are done for the nominal model with 1% reflectivity level (the results for models with 0.2% and 5% reflectivity are similar). As in the previous sections, the nominal model follows Subsection 5.1, and the background perturbation direction $\mathfrak{u}$ is that of Figure 17. Figure 27 shows the cost functions for FWI and MBTT at 7 Hz for different values of η. We also show by vertical lines the

exact R_G-attraction basin associated to zero tolerable error, and by horizontal lines the levels $1/2R_G^2$ corresponding to the tolerable error level associated to the exact Θ-attraction basin (cf. Table 4).

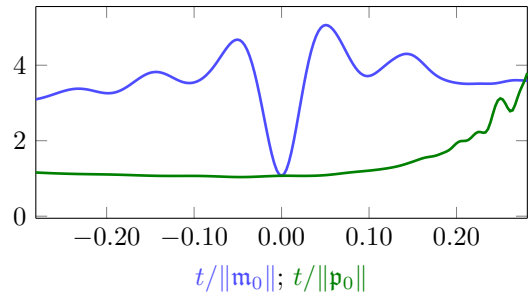
- The FWI and MBTT attraction basins are essentially independent of the error level η in the data. The size of the FWI R_G-attraction basins is slightly larger than the one anticipated by $\Delta^{\mathfrak{u}}_{\mathfrak{m}_0}$. When $\eta = 0$, $\boldsymbol{d} = \boldsymbol{d}_0$ is attainable, with a zero error level, and one sees that the first local minimum is just after the right limit of the exact R_G-attraction basin, hence consistent with the analysis of Section 3. The tolerable error level $R^{\mathfrak{u}}_{G,\mathfrak{p}_0}$ on $\boldsymbol{d}$ for the Θ-attraction basin is attained at $\eta = 0.1$.
- The local minima appear much earlier with FWI than with MBTT and the sizes of intervals where there is no local minimum is well anticipated by the estimates $\Delta^{\mathfrak{u}}$.
- In the numerical setting considered here, the MBTT approach appears to be robust with respect to data error levels much larger than the tolerable level, as we do not see local minimum appearing inside the attraction basin.



(a) Using $\eta = 0.1$.



(b) Using $\eta = 0$.



(c) Using $\eta = 1$.

Figure 27: Comparison of misfit functional $\mathcal{J}(\mathfrak{m}_0 + t\mathfrak{u})$ (blue) and $\mathsf{J}(\mathfrak{p}_0 + t\mathfrak{u}, \mathfrak{s}_0)$ (green) at 7 Hz, for different data $\boldsymbol{d}$ as defined in (57) for a nominal model with a reflectivity level $\beta = 1\%$. We also indicate by vertical dotted lines the size of the exact R_G-attraction basins, and by horizontal dotted lines the maximal tolerable errors $R^{\mathfrak{u}}_{G,m_0}$ associated to the exact Θ-attraction basins, taken from Table 4.

5.8 Strategy for MBTT waveform inversion

The numerical experiments made in this section lead to the following recommendations, for the implementation of the MBTT reformulation of FWI with a full wave equation, in order to obtain large attraction basins for the background reconstruction:

- the background space for $\mathfrak{p}$ has to be chosen smooth enough and the reflectivity level $\beta = \|\mathfrak{r}\|/\|\mathfrak{p}\|$ large enough so that the energy backscattered by $\mathfrak{p}$ is negligible compared to that backscattered by $\mathfrak{r}$;
- the reflectivity level β has to be chosen small enough for the importance of multiples to remain small compared to primary reflections in the synthetics.

This can be achieved by using an adapted parametrization for the background $\mathfrak{p}$, e.g. spline functions as employed in [22], and by tuning the reflectivity level β by adjusting the migration weight $\mathcal{W}$ in (52). We remind that the relevance of the chosen settings can be checked before performing any iterative optimization algorithm, by computing the size of the Θ and R_G-attraction basins, and the associated tolerable error.

- The attraction basins tend to be larger in the direction of slowly increasing (with depth) backgrounds, which suggests it is better to underestimate the initial slope of the background velocity with respect to depth.

The dependency of the size of attraction basins with the reflectivity level suggests the following organization for MBTT waveform inversion:

1. scale down the given seismic data so that they can be resimulated by a reflectivity level β of circa 0.2-1%, which amounts to bring the forward model close to the Born approximation.
2. Perform alternative minimization with respect to $\mathfrak{s}$ and $\mathfrak{p}$, starting with $\mathfrak{s}$ first.
3. Rescale (up) the processed seismic data in order to get closer to the original given data, and return to step 2.
4. Continue until processed and given original data coincide.

Note that all the difficulties related to the choice of the background and the reflectivity level disappear if one uses a linearized model (Born approximation) as forward modeling.

6 Extension for other applications of least squares minimization

The quantities that we have defined above only depend on the least-squares minimization problem we are considering. In particular, we have used it to compare the performance of methods in the context of the acoustic inverse wave problem in seismic. In addition, it is straightforward to apply the computational estimates towards different situations, which do not involve the Helmholtz equation, or the geophysical discrete setup. In this section, other situations are experimented: the inverse problem associated with elastic medium reconstruction, and problems with different kind of domain boundary conditions.

6.1 Elastic medium reconstruction

In the case of isotropic elastic propagation, the medium is characterized by three physical models: the Lamé parameters λ and μ, and the density ρ. The propagation of waves in such media is expressed in terms of the vectorial displacement field $\boldsymbol{u}$, see (1). More precisely, for elastic isotropic media, the equation of wave propagation writes as (where we omit the space dependency for clarity)

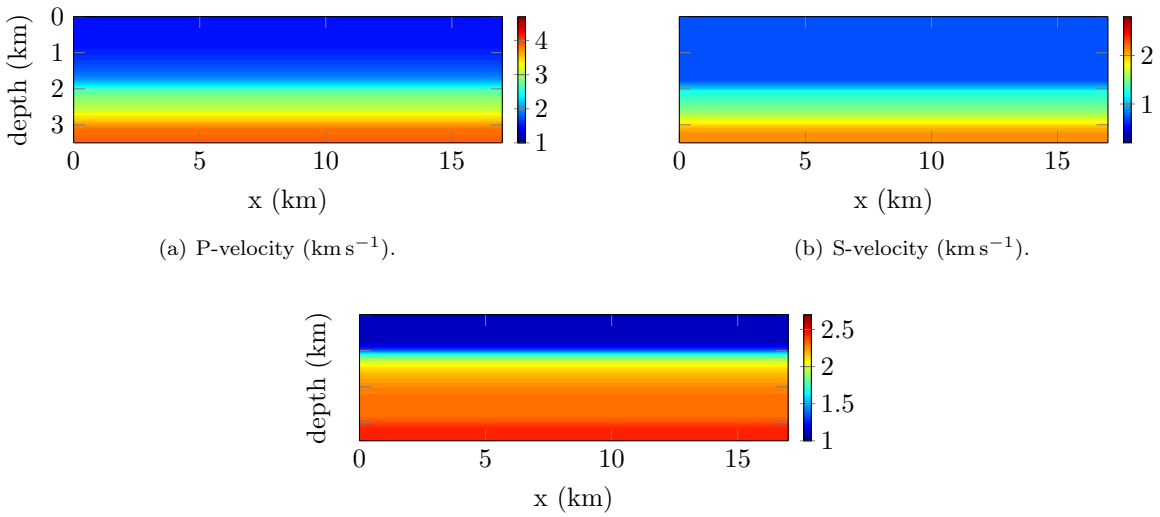
$$-\rho\omega^2\boldsymbol{u} - \nabla\Big(\lambda\nabla\cdot\boldsymbol{u}\Big) - \nabla\cdot\left(\mu\Big(\nabla\boldsymbol{u} + \big(\nabla\boldsymbol{u}\big)^T\Big)\right) = \boldsymbol{g}, \qquad \text{in } \Omega. \tag{58}$$

In this context the unknown m is composed of the three parameters, $m = \{\lambda,\ \mu,\ \rho\}$ (which are space dependent). The forward problem is accordingly adapted and we write

$$\mathcal{F}_{\text{elastic}}(m) = \boldsymbol{u}(\mathbf{x})|_{\Sigma}, \tag{59}$$

where we omit the source and frequency dependencies for clarity. Hence, we have considered that the forward problem is given by the measurements of displacement in all directions[2].

We consider the models m_0 to be smooth backgrounds, similarly to the previous acoustic case. They are illustrated in Figure 28 and represent a subsurface of size 17×3.5 km.



(a) P-velocity ($\mathrm{km\,s^{-1}}$).

(b) S-velocity ($\mathrm{km\,s^{-1}}$).

(c) density (10^3 $\mathrm{kg\,m^{-2}}$).

Figure 28: Elastic smooth model of size 17×3.5 km used for m_0 in the estimates. The relations between the velocities and Lamé parameters are $c_p = \sqrt{(\lambda + 2\mu)/\rho}$ and $c_s = \sqrt{\mu/\rho}$.

The geophysical acquisition is composed of 19 sources and 168 receivers for each source, both being located at the surface. In this case, the derivatives have to be computed with respect to the three parameters. For the directions, we consider the elastic Marmousi2 medium and extract the different reflectors. The directions for the Lamé parameters and the density are presented in Figure 29 where the amplitude has been chosen such that $\|\mathfrak{u}_m\| = 1$. Because of the difference of magnitude between the different quantities, we actually employ a scaling parameter to obtain dimensionless coefficients such that the parameters of interest are $(\lambda/\lambda_0, \mu/\mu_0, \rho/\rho_0)$, where λ_0, μ_0 and ρ_0 are scalar coefficients given by the maximal values of the respective quantities. In Figure 30, we show the evolution of the estimates $\Delta^{\mathfrak{u}}_{m_0}$ and $R^{\mathfrak{u}}_{m_0}$ with frequency.

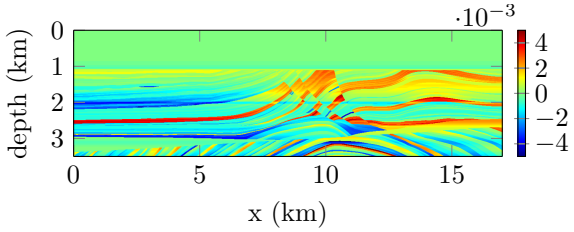
We see that the estimates for the size of the interval associated with the elastic wave equation, Figure 30(a) behave similarly to the acoustic case: the size of the attraction basin decreases with frequency. However, the maximal distance between the data and the attainable set, given Figure 30(b) appears now to be decreasing with frequency, contrary to the acoustic situation. It would be interesting to pursue the analysis in the elastic framework, in particular to identify how the different parameters are intertwined and how they should be taken.

6.2 Extension to different boundary conditions

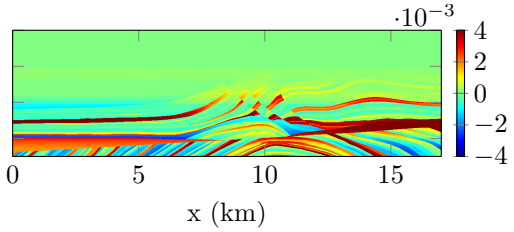
The estimates we have derived can also be employed for other situations than the inverse wave problem affiliated with a geophysical setup. Let us briefly illustrate the consideration of alternative boundary conditions for the domain of interest. We acknowledge two situations, sketched in Figure 31:

1. completely absorbing medium where the upper free surface is replaced by a perfectly matched layer,

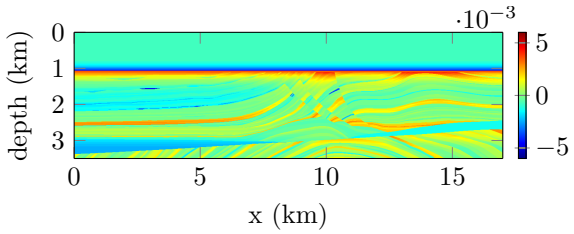
[2]We have also tried experiments restricting the available data to the vertical direction (i.e., u_z), which may be more relevant in seismic, and the results are the same. Hence, we only present the multi-direction data for simplicity.



(a) direction for the Lamé parameter λ (Pa)

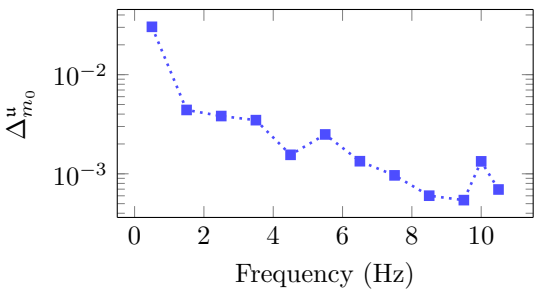


(b) direction for the Lamé parameter μ (Pa)

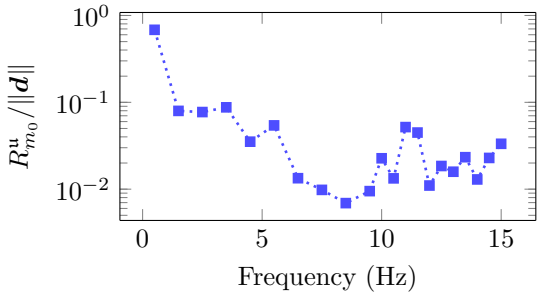


(c) direction for the density (10^3 kg m^{-3})

Figure 29: The elastic directions extracted from the Marmousi model. The amplitude is accordingly selected to give $\|\mathfrak{u}\| = 1$ for all coefficients (λ, μ and ρ).



(a) Evolution of $\Delta^{\mathfrak{u}}_{m_0}$ computed with local Θ-estimate.



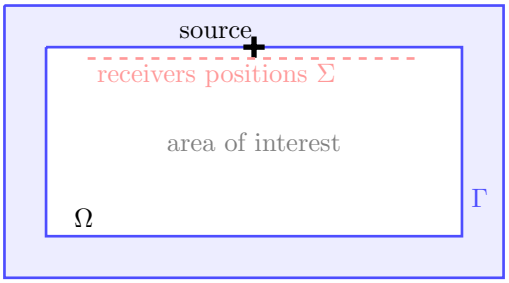
(b) Evolution of the maximal distance between the data and the attainable set, $R^{\mathfrak{u}}_{m_0}$.

Figure 30: Evolution of $\Delta^{\mathfrak{u}}_{m_0}$ and $R^{\mathfrak{u}}_{m_0}$ with frequency for elastic isotropic medium. The initial models m_0 are represented with smooth elastic background of Figure 28, the direction is given by the Marmousi structures of Figure 29. The computation of the directional derivatives uses scaled quantities: $(\lambda/\lambda_0, \mu/\mu_0, \rho/\rho_0)$.

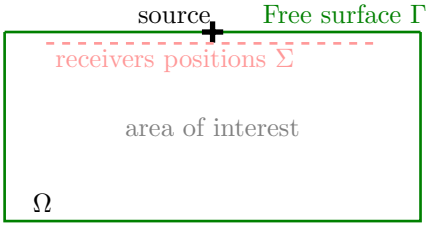
2. Dirichlet boundary condition set to zeros for all boundaries, generating reflection from all sides. This situation is particularly appropriate in the context of medical imaging and Electrical Impedance Tomography (EIT).

We compute the estimates according to the new situations and show the evolution of $\Delta^{\mathfrak{u}}_{m_0}$ (obtained with local Θ-estimate) and $R^{\mathfrak{u}}_{m_0}$ in Figures 32 and 33 respectively. The initial model is given by the smooth velocity background of Figure 3 and the directions are those introduced in Figure 4. Namely, we follow the setup of Subsection 4.1, and only the boundary conditions are changed.

The size of the attraction basin, Figure 32, decreases with increasing frequencies independently of the situation. However, in the case of Dirichlet boundaries on all sides, the behavior is much more distorted and reveals a much higher magnitude in the evolution. Namely it has four order of magnitude while only two for the other situation. Also in this case (Dirichlet boundary conditions), there does not seem to be any difference between the two directions, and the large obstacle direction, $\mathfrak{u}_3$, behaves similarly as the Marmousi ones $\mathfrak{u}_m$. Regarding the maximal distance to the attainable set

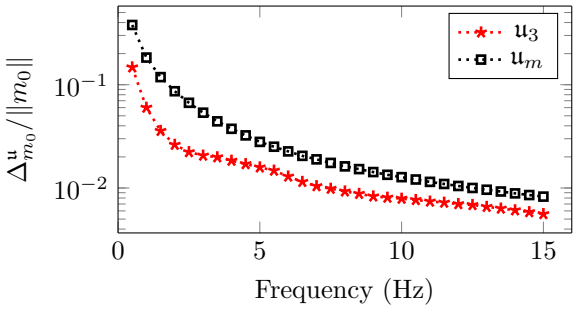


(a) Computational domain using PML for all boundaries.

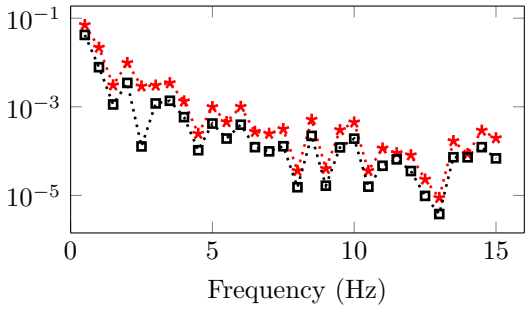


(b) Computational domain using Dirichlet boundary conditions only.

Figure 31: Illustration of different boundary conditions for the computational domain used to estimate the quantity $\Delta^{\mathrm{u}}_{m_0}$.

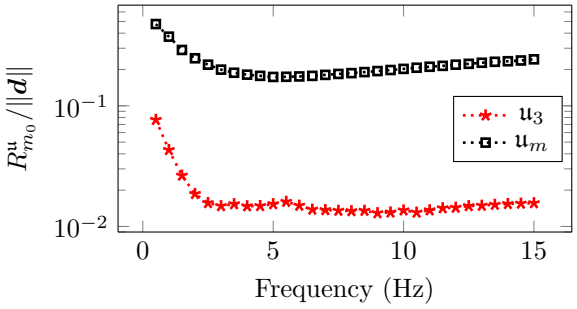


(a) Absorbing medium where all boundaries are treated as PML.

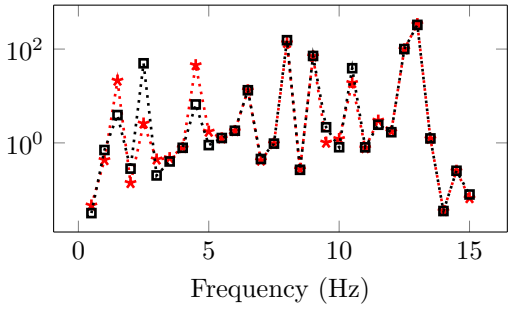


(b) Medium where all boundaries have a Dirichlet condition set to 0.

Figure 32: Evolution of $\Delta^{\mathrm{u}}_{m_0}$ computed with the local Θ-estimate with frequency for different kinds of boundary condition problems. We use m_0 as the smooth background model of Figure 3; the directions are given Figure 4.



(a) Absorbing medium where all boundaries are treated as PML.



(b) Medium where all boundaries have a Dirichlet condition set to 0.

Figure 33: Evolution of the maximal distance between the data and the attainable set, quantity $R^{\mathrm{u}}_{m_0}$, with frequency for different kinds of boundary condition problems. We use m_0 as the smooth background model of Figure 3; the directions are given Figure 4.

(largest tolerable eroor on data), Figure 33, we observe a decrease of the distance for low frequencies and then stabilization for the absorbing medium. This medium (fully absorbing) is strongly influenced by the geometry of the direction and the medium with Dirichlet boundary condition renders a totally unsettled evolution. We can intuit that the patterns obtained with the Dirichlet boundary conditions are due to the multiple reflections introduced at the boundaries, which strongly affect the problem in

terms of convergence and sensitivity to noise.

7 Conclusion

We have presented a theoretical and numerical toolbox for the apriori analysis of the optimizability of nonlinear least-squares minimization problems by local algorithms. Such problems arise in seismic Full Waveform Inversion (FWI), where the misfit function tends to exhibit local minima in the directions associated with low spatial frequencies perturbation of the background velocity. But the methodology is totally applicable for other least squares minimization problems.

We have defined *attraction basins* around a nominal model and the associated *tolerable error* level such that *for any data below tolerable error*, one is sure that the data misfit has *a single local - and hence global - minimum* over the basin. These quantities depend only on the forward map to be inverted, so one can obtain apriori information on the absence of local minima without having to experiment with the misfit function for different data. We have characterized attraction basins by the notion of global radius of curvature in the data space, and given formulas and algorithms for their numerical approximate and exact determination.

We have first applied this tool to the classical FWI problem for a wave equation in the time-frequency domain, and computed directional attraction basins which give quantitative insight into optimizability of FWI:

1. unsurprisingly, the size of attraction basins becomes smaller when frequency increases. In our tests, it is only of $\pm 2\%$ of the norm of the background at 4 Hz, and $\pm 1\%$ at 7 Hz,
2. the use of complex frequencies including damping enlarges the attraction basin,
3. the largest attraction basins obtained for low (or complex) frequency usually go with smaller tolerable error levels. Therefore, low and complex frequencies suffer more from noise.
4. solving the problem for a sequence of increasing single frequency data is better in terms of attraction basin than using an increasing sequence of frequency-bandwidth data,
5. salt domes produce smaller attraction basins, and are harder to recover.

These findings confirm the difficulty inherent to FWI, which requires unrealistically low frequency data for the determination of the full velocity model (Appendix C)

In order to see if these limitations could be overcome, we have then applied our tool to the MBTT approach, which uses a background/data space reflectivity reparameterization of the velocity model. Our findings are:

6. in a direction of background perturbation, the largest attraction basin for MBTT shows an increase in size by a factor four to fifteen compared to FWI, and is only barely reduced when frequency increases.
7. Correlatively, the tolerable error for MBTT is divided approximately by 10 compared to FWI. Nonetheless, the comparison of misfit functions has shown that local minimum does not appear when the tolerable error is exceeded.
8. In the direction of data-space reflectivity perturbations, where the forward map is linear up to the multiple reflections, the attraction basin is very large, more than ± 20 times the norm of the nominal data-space reflectivity.
9. Critical parameters for an efficient implementation of MBTT are the background smoothness and the reflectivity level.

This provides a strong incentive for the use of the MBTT decomposition to alleviate the low frequency requirement of FWI, despite a larger computational burden. For future work, we plan to implement the MBTT method with these settings to investigate further the efficiency in the context of iterative minimization with velocity model reconstruction.

A Adjoint state using complex variables

The Full Waveform Inversion (FWI) method aims the reconstruction of the subsurface medium parameters by an iterative minimization of the cost function defined as the difference between simulation and observations. We follow the standard least squares formulation of (13), and consider the Helmholtz equation (2) to write

$$\mathcal{J}(m) = \frac{1}{2}\sum_{\omega}\sum_{s}\|\mathcal{F}_{\omega}^{(s)}(m) - d^{(s)}\|_2^2 = \frac{1}{2}\sum_{\omega}\sum_{s}\|\mathcal{R}p_{\omega}^{(s)} - d_{\omega}^{(s)}\|_2^2, \tag{60}$$

where the forward problem is written with the restriction operator $\mathcal{R}$, and we use the index (s) for the sources. For the minimization, one needs to obtain the gradient of the cost function, which is usually obtained using adjoint state method, see [47] for a review of the method in geophysical application. We specify now the computations for complex-valued fields, omitting the source and frequency sums for clarity, and write

$$\mathcal{J}(m) = \frac{1}{2}\|\mathcal{F}(m) - \boldsymbol{d}\|_2^2 = \frac{1}{2}\|\mathcal{R}p - \boldsymbol{d}\|_2^2. \tag{61}$$

In the frequency domain, p is complex, which requires some precaution for the application of the adjoint state method. In particular, the functional

$$\mathrm{J}(m,p) = \frac{1}{2}\|\mathcal{R}p(m) - \boldsymbol{d}\|_2^2 \tag{62}$$

is not analytic (holomorphic) with respect to the field p. A workaround is relatively standard, see for example [14, 38, 36], with elements of complex calculus based on Wirtinger calculus. We believe it is important to mention this aspect which is too often disregarded in seismic applications and hereby present the steps involved.

A.1 Complex derivation

The derivation of complex functional is conducted by taking independently the complex variable and its conjugate, respectively z and $\overline{z}$, with $z = x + \mathrm{i}y$.

Theorem 1. *[14, Theorem 1] Let $g : \mathbb{C}\times\mathbb{C} \to \mathbb{C}$ be a function of a complex number z and its conjugate $\overline{z}$ and let g be analytic with respect to each variable (z and $\overline{z}$) independently. Let $h : \mathbb{R}\times\mathbb{R} \to \mathbb{C}$ be the function of the real variables x and y such that $g(z,\overline{z}) = h(x,y)$ where $z = x + \mathrm{i}y$. Then the partial derivative $\partial_z g$ (treating $\overline{z}$ as a constant) gives the same result as $(\partial_x h - \mathrm{i}\partial_y h)/2$. Similarly, $\partial_{\overline{z}} g$ is equivalent to $(\partial_x h + \mathrm{i}\partial_y h)/2$.*

Corollary 1. *Following the statement of Theorem 1, we have*

$$\overline{\frac{\partial g}{\partial \overline{z}}} = \frac{\partial g}{\partial z}. \tag{63}$$

Proof. By direct application of Theorem 1,

$$\overline{\frac{\partial g}{\partial \overline{z}}} = \frac{1}{2}\left(\overline{\frac{\partial f}{\partial x} + \mathrm{i}\frac{\partial f}{\partial y}}\right) = \frac{1}{2}\left(\frac{\partial f}{\partial x} - \mathrm{i}\frac{\partial f}{\partial y}\right) = \frac{\partial g}{\partial z}. \tag{64}$$

□

We straightforwardly apply the theorem to the misfit function where we consider $p := z = x + \mathrm{i}y$.

$$\mathrm{J} \;:\; (x,y) \quad \to \quad \frac{1}{2}\|\mathcal{R}(x+\mathrm{i}y) - d\|^2, \tag{65}$$

where x, y and d can be assimilated with vectors in the discrete setting. Then by deriving independently with respect to x and y we obtain

$$\begin{cases} \dfrac{\partial \mathrm{J}}{\partial x} = \dfrac{1}{2}\big[\mathcal{R}^*(\mathcal{R}(p) - d\,)\big]^T + \dfrac{1}{2}(\mathcal{R}(p) - d\,)^*\mathcal{R}, \\ \dfrac{\partial \mathrm{J}}{\partial y} = -\dfrac{\mathrm{i}}{2}\big[\mathcal{R}^*(\mathcal{R}(p) - d\,)\big]^T + \dfrac{\mathrm{i}}{2}(\mathcal{R}(p) - d\,)^*\mathcal{R}. \end{cases} \tag{66}$$

We can further deduce the derivative of J with respect to p and $\overline{p}$, where they are considered independent such that $\mathrm{J} = \mathrm{J}(p, \overline{p})$,

$$\begin{cases} \dfrac{\partial \mathrm{J}}{\partial p} = \dfrac{1}{2}\big[\mathcal{R}^*(\mathcal{R}(p) - d\,)\big]^T & = \dfrac{1}{2}(\mathcal{R}(p) - d\,)^T\mathcal{R}, \\ \dfrac{\partial \mathrm{J}}{\partial \overline{p}} = \dfrac{1}{2}(\mathcal{R}(p) - d\,)^*\mathcal{R} & = \dfrac{1}{2}(\mathcal{R}(\overline{p}) - \overline{d}\,)^T\mathcal{R} \end{cases} \tag{67}$$

The following theorems give the framework of what can be seen as the chain rule for complex derivation.

Theorem 2. *Consider the complex-valued function f of a real parameter m and the real-valued functions g_1 and g_2 such that $f(m) = g_1(z(m), \overline{z}(m)) + \mathrm{i}g_2(z(m), \overline{z}(m))$. The derivative with respect to the real parameter m is defined by*

$$\frac{\partial f}{\partial m} = \frac{\partial g}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}. \tag{68}$$

Proof. From the definition of f we have

$$\begin{aligned} \frac{\partial f}{\partial m} &= \frac{\partial g_1(z(m), \overline{z}(m))}{\partial m} + \mathrm{i}\frac{\partial g_2(z(m), \overline{z}(m))}{\partial m} \\ &= \frac{\partial g_1}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g_1}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} + \mathrm{i}\frac{\partial g_2}{\partial z}\frac{\partial z}{\partial m} + \mathrm{i}\frac{\partial g_2}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} \\ &= \frac{\partial (g_1 + \mathrm{i}g_2)}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial (g_1 + \mathrm{i}g_2)}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} \\ &= \frac{\partial g}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} \end{aligned} \tag{69}$$

□

Theorem 3. *Consider the real-valued functions f and g defined by $f(m) = g(z(m), \overline{z}(m))$. From Theorem 2 we have*

$$\frac{\partial f}{\partial m} = 2\,\mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right) = 2\,\mathrm{Re}\left(\frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}\right). \tag{70}$$

Proof. Direct application of Theorem 2 gives

$$\frac{\partial f}{\partial m} = \mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}\right). \tag{71}$$

We use Theorem 1 and Corollary 1, and take $z(m) = x(m) + \mathrm{i}y(m)$ to have

$$\mathrm{Re}\left(\frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}\right) = \mathrm{Re}\left(\overline{\frac{\partial g}{\partial z}}\frac{\partial \overline{z}}{\partial m}\right) = \mathrm{Re}\,\overline{\left(\frac{\partial g}{\partial z}\overline{\frac{\partial \overline{z}}{\partial m}}\right)} = \mathrm{Re}\left(\frac{\partial g}{\partial z}\overline{\frac{\partial \overline{z}}{\partial m}}\right), \tag{72}$$

where

$$\overline{\frac{\partial \overline{z}}{\partial m}} = \overline{\frac{\partial (x - \mathrm{i}y)}{\partial m}} = \overline{\left(\frac{\partial x}{\partial m} - \mathrm{i}\frac{\partial y}{\partial m}\right)} = \frac{\partial x}{\partial m} + \mathrm{i}\frac{\partial y}{\partial m} = \frac{\partial z}{\partial m}. \tag{73}$$

We inject in (71) to obtain

$$\frac{\partial f}{\partial m} = \mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right) + \mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right) = 2\,\mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right). \tag{74}$$

The alternative expression is obtained similarly but by replacing $\partial_z g$ in (71), instead of $\partial_{\overline{z}} g$. □

Application of Theorem 3 gives the gradient of the cost function with respect to m,

$$\begin{aligned} \nabla_m \mathrm{J} = \frac{\partial}{\partial m}\Big(\mathrm{J}(m,p)\Big)^T &= 2\,\mathrm{Re}\left(\frac{\partial \mathrm{J}}{\partial p}\frac{\partial p}{\partial m}\right)^T = \mathrm{Re}\left((\mathcal{R}(p) - d\,)^*\mathcal{R}\frac{\partial p}{\partial m}\right)^T \\ &= \mathrm{Re}\left(\left(\frac{\partial p}{\partial m}\right)^*\mathcal{R}^*(\mathcal{R}(p) - d\,)\right), \end{aligned} \tag{75}$$

where ^{T} stands for the transposed.

A.2 Adjoint state method

In order to avoid the computation of the Jacobian, the gradient is computed with the first order adjoint state method. It has been introduced in the work of [39], and implemented by [18] for the computation of a functional gradient. The formulation for the elastic wave problem has been carried out by [62, 63]. It is a relatively standard techniques nowadays, e.g. [34], see [47] for a review in geophysical situations. Yet, the complex variable specification is less common in seismic literature. In order to compute the derivative $\nabla\mathcal{J}$, we formulate the minimization problem (omitting the space dependency)

$$\min_{m\in\mathcal{M}} \mathcal{J}(m) = \mathrm{J}(m,p) \qquad \text{subject to } \mathcal{A}p = h, \tag{76}$$

where we introduce the wave operator $\mathcal{A}$, which corresponds to the Helmholtz equation in the acoustic case,

$$\mathcal{A} := \big(-\omega^2 c^{-2} - \Delta\big) = \big(-\omega^2 m - \Delta\big). \tag{77}$$

We consider one single source for now, to clarify the indexes, we shall later reintroduce the sum over the sources by linearity, cf. (85). The problem is recast into a formulation with Lagrangian to define

$$\mathcal{L}(m,\widehat{p},\widehat{\gamma}) = \mathrm{J}(m,\widehat{p}) + \big\langle \mathcal{A}\widehat{p} - h, \widehat{\gamma}\big\rangle, \tag{78}$$

where $\langle\cdot,\cdot\rangle$ stands for the complex inner product in L^2 such that $\langle v, w\rangle = v^* w$, with v^* the adjoint. By taking p solution of $\mathcal{A}p = h$, we have that $\nabla_m\mathcal{L}(m,p,\widehat{\gamma}) = \nabla_m\mathcal{J}(m)$. Furthermore by application of complex derivation Theorem 2, we have

$$\frac{\partial}{\partial m}\bigg(\mathcal{L}(m,p,\widehat{\gamma})\bigg) = \mathrm{Re}\left(\frac{\partial\mathcal{L}}{\partial m} + \frac{\partial\mathcal{L}}{\partial\overline{p}}\frac{\partial\overline{p}}{\partial m} + \frac{\partial\mathcal{L}}{\partial p}\frac{\partial p}{\partial m}\right). \tag{79}$$

By making use of the Corollary 1, we have

$$\frac{\partial}{\partial m}\bigg(\mathcal{L}(m,p,\widehat{\gamma})\bigg) = \mathrm{Re}\left(\frac{\partial\mathcal{L}}{\partial m} + \Big(\frac{\partial\mathcal{L}}{\partial\overline{p}} + \frac{\partial\mathcal{L}}{\partial p}\Big)\frac{\partial p}{\partial m}\right). \tag{80}$$

The adjoint state γ is now selected such that

$$\mathrm{Re}\left(\frac{\partial\mathcal{L}}{\partial\overline{p}} + \frac{\partial\mathcal{L}}{\partial p}\right) = 0, \tag{81}$$

which gives,

$$\mathrm{Re}\left(\frac{\partial\mathrm{J}}{\partial p} + \frac{\partial\mathrm{J}}{\partial\overline{p}} + \mathcal{A}^*\gamma\right) = 0. \tag{82}$$

We now incorporate (67), and the adjoint state γ solves the problem

$$\mathcal{A}^*\gamma = -\mathcal{R}^*\big(\mathcal{R}(p) - d\big). \tag{83}$$

Using this formulation for γ, the gradient reduces to

$$\nabla_m\mathcal{J} = \mathrm{Re}\left(\langle\partial_m\mathcal{A}p, \gamma\rangle\right)^T. \tag{84}$$

We can reintroduce the sum over the sources, which gives

$$\nabla_m\mathcal{J} = \sum_s \mathrm{Re}\left(\langle(\partial_m\mathcal{A})p^{(s)}, \gamma^{(s)}\rangle\right)^T, \quad \text{where } \gamma^{(s)} \text{ solves } \mathcal{A}^*\gamma^{(s)} = -\mathcal{R}^*\big(\mathcal{R}(p^{(s)}) - d^{(s)}\big). \tag{85}$$

A.3 Directional derivative computation

For the computation of the directional Fréchet derivative, required for our estimates, the technique is quite different. Note that from the forward operator definition, the Fréchet derivative consists in computing the derivative of the pressure field and to apply the linear restriction $\mathcal{R}$. Thus we focus on the following in the computation of the directional derivative of the pressure field. Here we apply a model perturbation onto the wave equation, in a similar approach as the one prescribed in [28]. Let us consider a model perturbation $\mathfrak{u}$ for the Helmholtz equation (2), where the parameter m is identified with c^{-2},

$$-\omega^2\big(m(\mathbf{x})+\mathfrak{u}(\mathbf{x})\big)p(\mathbf{x})-\Delta p(\mathbf{x})=h(\mathbf{x}). \tag{86}$$

The wave field solution is decomposed into

$$p:=p(m+\mathfrak{u})=p(m)+p_{\mathfrak{u}}=p_0+p_{\mathfrak{u}}, \tag{87}$$

where p_0 is solution to the original (non-perturbed) equation (2) and the space dependency is omitted for clarity. Developing (86) gives

$$\begin{aligned}\big(-\omega^2 m-\Delta\big)p-\omega^2\mathfrak{u}p&=h\\ h+\big(-\omega^2 m-\Delta\big)p_{\mathfrak{u}}-\omega^2\mathfrak{u}p&=h\\ \big(-\omega^2 m-\Delta\big)p_{\mathfrak{u}}&=\omega^2\mathfrak{u}p.\end{aligned} \tag{88}$$

Hence, $p_{\mathfrak{u}}$ is solution to the Helmholtz equation with some right-hand side depending on the total solution p. Assuming ω is not an eigenfrequency of the Helmholtz equation, we write

$$p_{\mathfrak{u}}=\big(-\omega^2 m-\Delta\big)^{-1}\big(\omega^2\mathfrak{u}p\big). \tag{89}$$

Consequently, using the definition $p=p_0+p_{\mathfrak{u}}$, we have

$$\begin{aligned}\Big(I-\big(-\omega^2 m-\Delta\big)^{-1}\mathfrak{u}\omega^2\Big)p&=p_0,\\ \Big(I-\big(-\omega^2 m-\Delta\big)^{-1}\mathfrak{u}\omega^2\Big)^{-1}p_0&=p,\end{aligned} \tag{90}$$

where I denotes the identity. Direct application of Neumann series provides

$$\begin{aligned}\Big(I-\big(-\omega^2 m-\Delta\big)^{-1}\mathfrak{u}\omega^2\Big)^{-1}p_0&=p_0+\sum_{k=1}^{\infty}\Big(\big(-\omega^2 m-\Delta\big)^{-1}\mathfrak{u}\omega^2\Big)^k p_0\\ &=p_0+\sum_{k=1}^{\infty}p_k.\end{aligned} \tag{91}$$

Using the linearity of the wave equation, we have that p_1 is solution to

$$\big(-\omega^2 m-\Delta\big)p_1=\mathfrak{u}\omega^2 p_0, \tag{92}$$

and repeatedly,

$$\big(-\omega^2 m-\Delta\big)p_j=\mathfrak{u}\omega^2 p_{j-1}. \tag{93}$$

By assuming the perturbation $\mathfrak{u}$ of the model to be small, the Taylor expansion allows us to express (with the notation of [17]),

$$\begin{aligned}p(m+\mathfrak{u})&=p(m)+\frac{\partial p(m)}{\partial m}(\mathfrak{u})+\frac{1}{2}\frac{\partial^2 p(m)}{\partial m^2}(\mathfrak{u},\mathfrak{u})+\dots\\ &=p_0+\frac{\partial p(m)}{\partial m}(\mathfrak{u})+\frac{1}{2}\frac{\partial^2 p(m)}{\partial m^2}(\mathfrak{u},\mathfrak{u})+\dots\end{aligned} \tag{94}$$

We can finally identify the directional derivative in terms of Neumann series and the directional derivative is expressed as solution to a forward problem, i.e. $\partial_m p(\mathfrak{u})=p_1$; $\partial_m^2 p(\mathfrak{u},\mathfrak{u})=2p_2$; etc. Note that obtaining the derivative from a forward problem can also be found in the context of elastic-fluid interaction in [4], where the derivation is conducted with respect to the Lamé parameters.

B MBTT (Data Space) model parameterization

In this appendix, we explicit how the computations are conducted in the case of MBTT model decomposition.

B.1 Expression of the reflectivity with MBTT

The reflectivity part in the MBTT model representation is given by, cf. (15),

$$\mathfrak{r} = \mathcal{W} D\mathcal{F}_0^* \mathfrak{s}. \tag{95}$$

The technique developed with the adjoint state method allows to compute of $\mathfrak{r}$ without explicitly using $D\mathcal{F}_0$. Indeed, the adjoint state method provides (identification from (75) and (85))

$$\begin{aligned} \sum_s \left(\langle (\partial_m \mathcal{A}) p^{(s)}, \gamma^{(s)} \rangle \right)^T &= \sum_s \left(\frac{\partial p^{(s)}}{\partial m} \right)^* \mathcal{R}^* (\mathcal{R}(p^{(s)}) - d^{(s)}) \\ \Rightarrow \sum_s \left(\big(\partial_m \mathcal{A}) p^{(s)} \big)^* \gamma^{(s)} \right)^T &= D\mathcal{F}^{(s)}(m)^* (\mathcal{R}(p^{(s)}) - d^{(s)}), \end{aligned} \tag{96}$$

where $D\mathcal{F}^{(s)}$ stands for the forward operator associated with source s. The fields $p^{(s)}$ and $\gamma^{(s)}$ solve respectively the forward and ajoint problems, see (2) and (85). Proceeding by analogy with the reasoning of Appendix A, it is straightforward to see that $\mathfrak{r}$ can be express as

$$\mathfrak{r} = \mathcal{W} \sum_s \left(\big((\partial_{\mathfrak{p}} \mathcal{A}_0) p_0^{(s)} \big)^* \gamma_0^{(s)} \right)^T, \tag{97}$$

where $\mathcal{A}_0$ is the Helmholtz operator with zero reflectivity (i.e., $m = \mathfrak{p}$),

$$\mathcal{A}_0 := \big(-\omega^2 \mathfrak{p} - \Delta \big); \tag{98}$$

$p_0^{(s)}$ solves the forward problem with $\mathcal{A}_0$ for the source s, and γ_0 solves for the source s,

$$\mathcal{A}_0^* \gamma_0^{(s)} = -\mathcal{R}^* \mathfrak{s}^{(s)}, \tag{99}$$

where $\mathfrak{s}$ writes as $\mathfrak{s} = \{\mathfrak{s}^{(1)} \dots \mathfrak{s}^{(n_{src})}\}$. The model representation (15) becomes

$$m(\mathfrak{p}, \mathfrak{s}) = \mathfrak{p} + \mathcal{W} \sum_s \left(\big((\partial_{\mathfrak{p}} \mathcal{A}_0) p_0^{(s)} \big)^* \gamma_0^{(s)} \right)^T, \tag{100}$$

Process summary In order to express the model and the gradient of the cost function in this case, the workflow is as follows

1. start from prescribed representation $\mathfrak{p}$ and $\mathfrak{s}$,
2. solve the Helmhotz equation with zero-reflectivity for all sources s, $\mathcal{A}_0 p_0^{(s)} = h^{(s)}$,
3. solve the adjoint problem to retrieve $\gamma_0^{(s)}$: $\mathcal{A}_0^* \gamma_0^{(s)} = -\mathcal{R}^* \mathfrak{s}^{(s)}$,
4. compute the model m with (100),
5. solve the Helmholtz equation for $p^{(s)}$: $\mathcal{A} p^{(s)} = h^{(s)}$,
6. solve the adjoint problem for $\gamma^{(s)}$: $\mathcal{A}^* \gamma^{(s)} = -\mathcal{R}^* \big(\mathcal{R}(p^{(s)}) - d^{(s)} \big)$.

Finally, the gradient of the cost function, with respect to $\mathfrak{p}$ or $\mathfrak{s}$ is expressed from the adjoint state method,

$$\begin{cases} \nabla_{\mathfrak{p}} \mathcal{J} = \displaystyle\sum_s \mathrm{Re} \left(\langle (\partial_{\mathfrak{p}} \mathcal{A}) p^{(s)}, \gamma^{(s)} \rangle \right)^T, \\ \nabla_{\mathfrak{s}} \mathcal{J} = \displaystyle\sum_s \mathrm{Re} \left(\langle (\partial_{\mathfrak{s}} \mathcal{A}) p^{(s)}, \gamma^{(s)} \rangle \right)^T. \end{cases} \tag{101}$$

B.2 Directional derivative computation

For the estimation of the size of the basin of attraction and of the radius of curvature, we need the directional derivative of the forward operator. Using Neumann series and Taylor expansion, we have given in Subsection A.3 their expression using the traditional model formulation. When the model is decomposed using MBTT, one can process similarly, it simply requires a few more steps. For clarity we only focus on the parameter $\mathfrak{p}$, we start with the chain rule,

$$\frac{\partial \mathsf{F}}{\partial \mathfrak{p}}(\mathfrak{u}_\mathfrak{p}) = \frac{\partial \mathsf{F}}{\partial m}\frac{\partial m}{\partial \mathfrak{p}}(\mathfrak{u}_\mathfrak{p}). \tag{102}$$

We derive from (100),

$$\left(\frac{\partial m}{\partial \mathfrak{p}}(\mathfrak{u}_\mathfrak{p})\right)^T = \mathfrak{u}_\mathfrak{p} + \mathcal{W}\sum_s \left((\partial_\mathfrak{p}\mathcal{A}_0)\partial_\mathfrak{p} p_0^{(s)} + (\partial^2_{\mathfrak{p}^2}\mathcal{A}_0)p_0^{(s)}\right)^*(\mathfrak{u}_\mathfrak{p})\gamma_0^{(s)} + \left((\partial_\mathfrak{p}\mathcal{A}_0)p_0^{(s)}\right)^*\partial_\mathfrak{p}\gamma_0^{(s)}(\mathfrak{u}_\mathfrak{p}). \tag{103}$$

The terms $(\partial_\mathfrak{p} p_0)(\mathfrak{u}_\mathfrak{p})$ and $(\partial_\mathfrak{p}\gamma_0)(\mathfrak{u}_\mathfrak{p})$ are computed using the same method as presented in Subsection A.3 (thus, each requires the resolution of the wave equation with specific right-hand side). After we have computed $\mathfrak{u}_{\widehat{m}} = (\partial_\mathfrak{p} m)(\mathfrak{u}_\mathfrak{p})$, we again use Subsection A.3 to compute the final $\partial_m \mathsf{F}(\mathfrak{u}_{\widehat{m}})$.

Eventually, one can process similarly for $\mathfrak{s}$, adapting the chain rule. Regarding the second order derivatives, it is analogous with one degree more of derivation in the chain rule. Computationally speaking, it only requires the resolution of additional forward problems.

C Influence of the geometry and frequency on FWI

We illustrate the influence of the subsurface geometry with an acoustic FWI experiment, where we demonstrate the effect we have depicted in Section 4, Figure 5:

1. a high contrast object complicates the process of reconstruction compared to a Marmousi-like subsurface, by reducing the size of the basin of attraction,
2. the low frequencies, by increasing the size of the basin of attraction, provide benefits on the reconstruction.

We design an experiment for a two-dimensional area of size 12.3×3 km, considering an acoustic medium with constant density set to 1000 kg m^{-2}. The wave propagation is governed by the Helmholtz equation (2), one parameter has to be recovered: the velocity c. The forward operator, see (10), provides pressure measurement at the receivers location. For the seismic acquisition, we take 135 sources and 306 receivers per source. We take synthetic data, avoiding noise, simply to illustrate our analysis.

C.1 Comparison between salt and Marmousi geometries

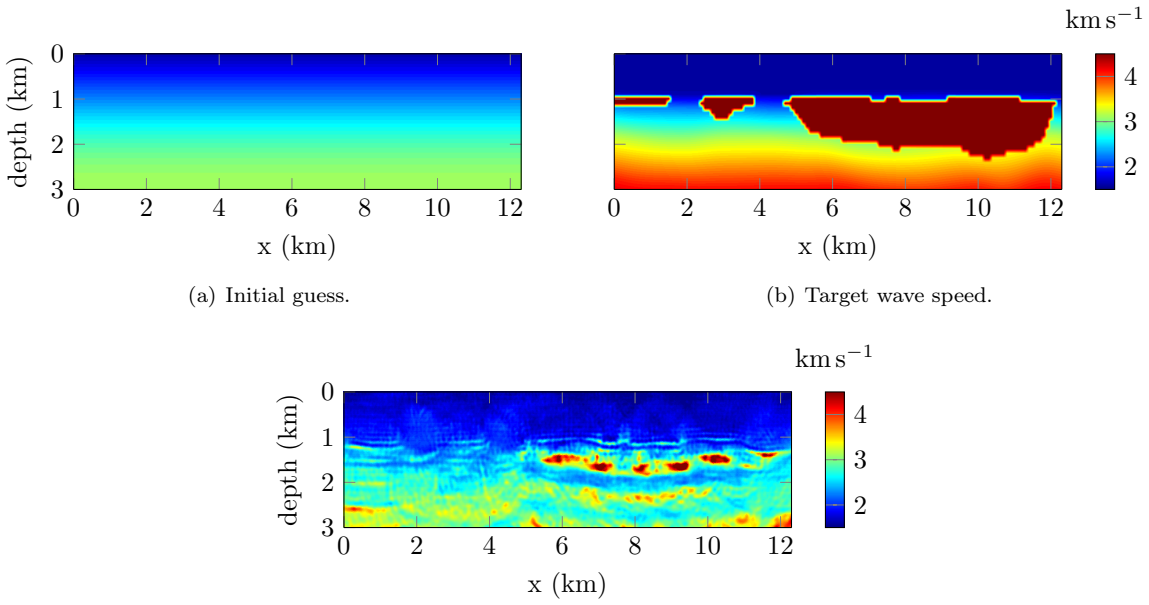
We compare the subsurface reconstruction using FWI depending on the geometry of the subsurface. We consider a medium involving a high contrasting object, and the Marmousi model. The iterative minimization is conducted following a nonlinear conjugate gradient algorithm (see [44]) and using frequencies from 1 to 10 Hz, with 1 Hz increment. We perform 20 iterations per frequencies, which are taken sequentially. The corresponding FWI algorithm is shown in Algorithm 1.

For the experiment related to the Marmousi medium, the 'true' model has been given in Figure 8(a) and the initial guess in Figure 3. Here we artificially modify the dimension so that the models are now of size 12.3×3 km. For the model encompassing a salt dome, the initial and target models are given in Figures 34(a) and 34(b) respectively.

In Figures 34(c) and 35, we compare the reconstructions from FWI regarding the contrasting object or the Marmousi model. Both situations use media of same size, with the same frequency progression (sequential from 1 to 10 Hz), and start from initial models where no information is known (one dimensional wave speed variation, with depth only). However, only the Marmousi case is able to accurately capture the subsurface structures. It demonstrates the results of Section 4 that, namely, high contrast objects, by reducing the size of the basin of attraction, are more complicated to reconstruct.

Initial parameters: initial model c_1 and the frequency list of ω_i, for $i = 1, \ldots, n_\omega$.
The frequencies are ordered such that $\omega_1 < \omega_2 < \ldots < \omega_{n_\omega}$, see Remark 4.
Frequency loop for $i \in \{1, \ldots, n_\omega\}$ **do**
  Optimization loop for $j \in \{1, \ldots, n_{iter}\}$ **do**
    set $k := (i-1)n_{\text{iter}} + j$
    - solve the wave equation at frequency ω_i with wave speed c_k;
    - compute the cost function by comparing the simulation and observation data;
    - compute the gradient of the cost function, using the adjoint-state method of Appendix A;
    - compute the search direction, s_k with the nonlinear conjugate gradient method (see [44]);
    - compute the step length α with line search method (see [44]);
    - update the wave speed $c_{k+1} = c_k - \alpha s_k$;
  end
end

Algorithm 1: FWI iterative minimization algorithm for the wave speed reconstruction.



(a) Initial guess.

(b) Target wave speed.

(c) FWI wave speed reconstruction using frequencies from 1 to 10 Hz.

Figure 34: 2D reconstruction of wave speed including salt domes with data covering frequencies from 1 to 10 Hz. The model is of size 12.3×3 km.

C.2 Low frequencies compensation

It has been shown in Figure 5 that low frequencies enlarge the size of the radius. To illustrate this effect, we reproduce the previous experiment but incorporating nine frequencies from 0.1 to 0.9 Hz (sequentially using 0.1 Hz increment). Then, we employ the set of frequencies from 1 to 10 Hz (sequentially using 1 Hz increment). In Figure 36, we show the reconstruction starting from these very low (unrealistic) frequencies.

The FWI algorithm benefits from this incorporation of low frequencies, and is now able to accurately retrieve the contrasting objects, as expected. Here we have used synthetic experiment to demonstrate the basin of attraction dependency on the geometry and frequency. Note that we could as well have used complex frequencies to increase the basin of attraction in early stages, cf. Subsection 4.5. However, for realistic application, low frequencies are usually unavailable (due to the noise)

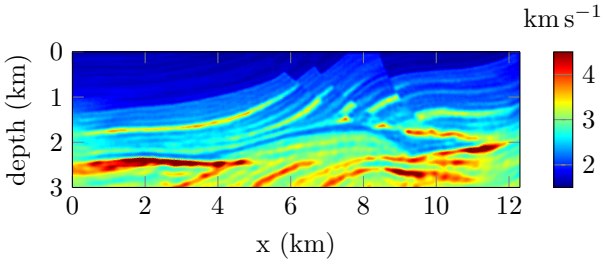


Figure 35: 2D reconstruction of the Marmousi wave speed with synthetic data covering frequencies from 1 to 10 Hz. The dimensions are artificially increased to the size 12.3×3 km compared to the true and starting media, given Figures 8(a) and 3 respectively.

and it is then impossible to reduce arbitrarily the initial frequency. As an alternative, techniques based on shape reconstruction could be very efficient in this context but, because the nature of the target is unknown at the beginning, it is hard to justify apriori.

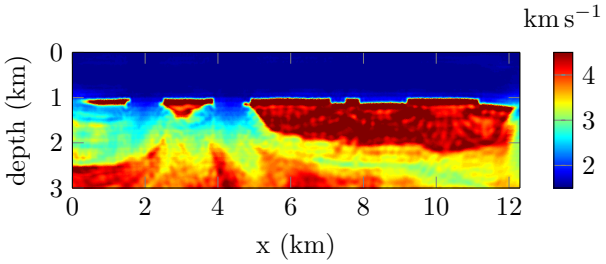


Figure 36: 2D reconstruction of wave speed including salt domes with data covering frequencies from 0.1 to 10 Hz. The initial model used for the algorithm is given in Figure 34(a), and the true model Figure 34(b). The algorithm takes full benefits of the low frequency content in the data because we use synthetic measurements without noise.

References

[1] V. Akcelik, G. Biros, and O. Ghattas, *Parallel multiscale gauss-newton-krylov methods for inverse wave propagation*, in Supercomputing, ACM/IEEE 2002 Conference, Nov 2002.

[2] G. Alessandrini, *Stable determination of conductivity by boundary measurements*, Appl. Anal., 27 (1988), pp. 153–172.

[3] G. Alessandrini and S. Vessella, *Lipschitz stability for the inverse conductivity problem*, Adv. in Appl. Math., 35 (2005), pp. 207–241.

[4] I. Azpiroz, H. Barucq, R. Djellouli, and H. Pham, *Characterization of partial derivatives with respect to material parameters in a fluid–solid interaction problem*, Journal of Mathematical Analysis and Applications, 465 (2018), pp. 903–927.

[5] A. Bamberger, G. Chavent, and P. Lailly, *Une application de la théorie du contrôle à un problème inverse de sismique*, Annales de Géophysique, 33 (1977), pp. 183–200.

[6] ——, *About the stability of the inverse problem in the 1-d wave equation*, Journal of Applied Mathematics and Optimisation, 5 (1979), pp. 1–47.

[7] G. Bao and P. Li, *Inverse medium scattering for the helmholtz equation at fixed frequency*, Inverse Problems, 21 (2005), p. 1621.

[8] ——, *Numerical solution of an inverse medium scattering problem for Maxwell's equations at fixed frequency*, Journal of Computational Physics, 228 (2009), pp. 4638–4648.

[9] H. Barucq, F. Faucher, and H. Pham, *Localization of small obstacles from back-scattered data at limited incident angles with full-waveform inversion*, Journal of Computational Physics, 370 (2018), pp. 1–24.

[10] J.-P. Bérenger, *A perfectly matched layer for the absorption of electromagnetic waves*, Journal of Computational Physics, 114 (1994), pp. 185 – 200.

[11] E. Beretta, M. V. de Hoop, F. Faucher, and O. Scherzer, *Inverse boundary value problem for the helmholtz equation: quantitative conditional lipschitz stability estimates*, SIAM Journal on Mathematical Analysis, 48 (2016), pp. 3962–3983.

[12] E. Beretta, M. V. de Hoop, and L. Qiu, *Lipschitz stability of an inverse boundary value problem for a schrödinger type equation*, SIAM Journal on Mathematical Analysis, 45 (2013), pp. 679–699.

[13] E. Bozdağ, J. Trampert, and J. Tromp, *Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements*, Geophysical Journal International, 185 (2011), pp. 845–870.

[14] D. Brandwood, *A complex gradient operator and its application in adaptive array theory*, in IEE Proceedings F-Communications, Radar and Signal Processing, vol. 130, IET, 1983, pp. 11–16.

[15] A. J. Brenders, U. Albertin, and J. Mika, *Comparison of 3d time-and frequency-domain waveform inversion: benefits and insights of a broadband, discrete-frequency strategy*, in SEG Technical Program Expanded Abstracts 2012, Society of Exploration Geophysicists, 2012, pp. 1–5.

[16] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, *Multiscale seismic waveform inversion*, Geophysics, 60 (1995), pp. 1457–1473.

[17] H. Cartan, *Differential calculus*, vol. 1, Hermann, 1971.

[18] G. Chavent, *Identification of functional parameters in partial differential equations*, in Identification of Parameters in Distributed Systems, R. E. Goodson and M. Polis, eds., ASME, New York, 1974, pp. 31–48.

[19] G. Chavent, *Nonlinear least squares for inverse problems: theoretical foundations and step-by-step guide for applications*, Springer Science & Business Media, 2010.

[20] G. Chavent, *Data Space Reflectivity and the Migration based Travel Time approach to FWI*, in 79th EAGE Conference and Exhibition 2017-Workshops, 2017.

[21] G. Chavent and F. Clément, *Waveform inversion through MBTT formulation*, INRIA, 1992.

[22] G. Chavent, K. Gadylshin, and V. Tcheverda, *Reflection fwi in mbtt formulation*, in 77th EAGE Conference and Exhibition 2015, 2015.

[23] G. Chavent and K. Kunisch, *On weakly nonlinear inverse problems*, SIAM Journal on Applied Mathematics, 56 (1996), pp. 542–572.

[24] F. Clément, G. Chavent, and S. Gómez, *Migration-based traveltime waveform inversion of 2-d simple structures: A synthetic example*, Geophysics, 66 (2001), pp. 845–860.

[25] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, Applied mathematical sciences, Springer, New York, second edition ed., 1998.

[26] E. Crase, A. Pica, M. Noble, J. McDonald, and A. Tarantola, *Robust elastic nonlinear waveform inversion: Application to real data*, Geophysics, 55 (1990), pp. 527–538.

[27] M. V. de Hoop, L. Qiu, and O. Scherzer, *A convergence analysis of a multi-level projected steepest descent iteration for nonlinear inverse problems in banach spaces subject to stability constraints*, arXiv preprint arXiv:1206.3706, (2012).

[28] L. Demanet, *Waves and imaging*, March 2017. Lecture notes, Massachusetts Institute of Technology.

[29] S. C. Eisenstat and H. F. Walker, *Globally convergent inexact newton methods*, SIAM Journal on Optimization, 4 (1994), pp. 393–422.

[30] B. Engquist and A. Majda, *Absorbing boundary conditions for numerical simulation of waves*, Proceedings of the National Academy of Sciences, 74 (1977), pp. 1765–1766.

[31] F. Faucher, *Contributions to seismic full waveform inversion for time harmonic wave equations: stability estimates, convergence analysis, numerical experiments involving large scale optimization algorithms*, PhD thesis, Université de Pau et Pays de l'Ardour, 2017.

[32] O. Gauthier, J. Virieux, and A. Tarantola, *Two-dimensional nonlinear inversion of seismic waveforms; numerical results*, Geophysics, 51 (1986), pp. 1387–1403.

[33] C. B. James Martin, Lucas C. Wilcox and O. Ghattas, *A stochastic newton mcmc method for large-scale statistical inverse problems with application to seismic inversion*, SIAM Journal on Scientific Computing, 34 (2012), pp. A1460–A1487.

[34] M. Kern, *Numerical Methods for Inverse Problems*, John Wiley & Sons, 2016.

[35] A. Kirsch, *An introduction to the mathematical theory of inverse problems*, Applied mathematical sciences, Springer, New York, 1996.

[36] K. Kreutz-Delgado, *The complex gradient operator and the cr-calculus*, arXiv preprint arXiv:0906.4835, (2009).

[37] P. Lailly, *The seismic inverse problem as a sequence of before stack migrations*, in Conference on Inverse Scattering: Theory and Application, J. B. Bednar, ed., Society for Industrial and Applied Mathematics, 1983, pp. 206–220.

[38] H. Li and T. Adali, *Optimization in the complex domain for nonlinear adaptive filtering*, in Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on, IEEE, 2006, pp. 263–267.

[39] J. L. Lions and S. K. Mitter, *Optimal control of systems governed by partial differential equations*, vol. 1200, Springer Berlin, 1971.

[40] N. Mandache, *Exponential instability in an inverse problem for the Schrödinger equation*, Inverse Problems, 17 (2001), pp. 1435–1444.

[41] L. Métivier, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, *Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion*, Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, 205 (2016), pp. 345–377.

[42] L. Métivier, R. Brossier, J. Virieux, and S. Operto, *Full waveform inversion and the truncated newton method*, SIAM Journal on Scientific Computing, 35 (2013), pp. B401–B437.

[43] P. Mora, *Nonlinear two-dimensional elastic inversion of multioffset seismic data*, Geophysics, 52 (1987), pp. 1211–1228.

[44] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Sries in Operations Research, 2 ed., 2006.

[45] P. V. Petrov and G. A. Newman, *Three-dimensional inverse modelling of damped elastic wave propagation in the fourier domain*, Geophysical Journal International, 198 (2014), pp. 1599–1617.

[46] R. Plessix, G. Chavent, and Y. De Roeck, *A quantitative kirchhoff migration to esimate the 2d velocity distribution*, in 3rd Internat. Conf. on Mathematical and Numerical Aspects of Wave Propagation, 1995, pp. 704–712.

[47] R.-E. Plessix, *A review of the adjoint-state method for computing the gradient of a functional with geophysical applications*, Geophysical Journal International, 167 (2006), pp. 495–503.

[48] R. G. Pratt and N. R. Goulty, *Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data*, Geophysics, 56 (1991), pp. 208–224.

[49] R. G. Pratt, C. Shin, and G. Hick, *Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion*, Geophysical Journal International, 133 (1998), pp. 341–362.

[50] R. G. Pratt, Z.-M. Song, P. Williamson, and M. Warner, *Two-dimensional velocity models from wide-angle seismic data by wavefield inversion*, Geophysical Journal International, 124 (1996), pp. 323–340.

[51] R. G. Pratt and M. H. Worthington, *Inverse theory applied to multi-source cross-hole tomography.*, Geophysical Prospecting, 38 (1990), pp. 287–310.

[52] L. Qiu, J. Ramos-Martínez, A. Valenciano, Y. Yang, and B. Engquist, *Full-waveform inversion with an exponentially encoded optimal-transport norm*, in SEG Technical Program Expanded Abstracts 2017, Society of Exploration Geophysicists, 2017, pp. 1286–1290.

[53] C. Shin and Y. H. Cha, *Waveform inversion in the laplace domain*, Geophysical Journal International, 173 (2008), pp. 922–931.

[54] ——, *Waveform inversion in the laplace fourier domain*, Geophysical Journal International, 177 (2009), pp. 1067–1079.

[55] C. Shin, S. Jang, and D.-J. Min, *Improved amplitude preservation for prestack depth migration by inverse scattering theory*, Geophysical Prospecting, 49 (2001), pp. 592–606.

[56] C. Shin, N.-H. Koo, Y. H. Cha, and K.-P. Park, *Sequentially ordered single-frequency 2-d acoustic waveform inversion in the laplace–fourier domain*, Geophysical Journal International, 181 (2010), pp. 935–950.

[57] C. Shin and D.-J. Min, *Waveform inversion using a logarithmic wavefield*, Geophysics, 71 (2006), pp. R31–R42.

[58] C. Shin, S. Pyun, and J. B. Bednar, *Comparison of waveform inversion, part 1: conventional wavefield vs logarithmic wavefield*, Geophysical Prospecting, 55 (2007), pp. 449–464.

[59] L. Sirgue and R. G. Pratt, *Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies*, Geophysics, 69 (2004), pp. 231–248.

[60] W. Symes and J. J. Carazzone, *Velocity inversion by differential semblance optimization*, Geophysics, 56 (1991), pp. 654–663.

[61] A. Tarantola, *Inversion of seismic reflection data in the acoustic approximation*, Geophysics, 49 (1984), pp. 1259–1266.

[62] A. Tarantola, *Inversion of travel times and seismic waveforms*, in Seismic tomography, Springer, 1987, pp. 135–157.

[63] A. Tarantola, *Theoretical background for the inversion of seismic waveforms including elasticity and attenuation*, Pure and Applied Geophysics, 128 (1988), pp. 365–399.

[64] L. Thomsen, *Weak elastic anisotropy*, Geophysics, 51 (1986), pp. 1954–1966.

[65] I. Tsvankin, *Anisotropic parameters and p-wave velocity for orthorhombic media*, Geophysics, 62 (1997), pp. 1292 – 1309.

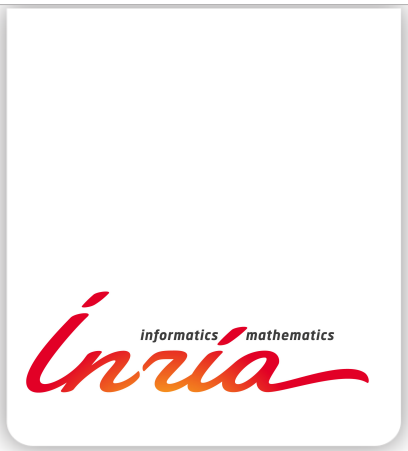
[66] E. Turkel and A. Yefet, *Absorbing pml boundary layers for wave-like equations*, Applied Numerical Mathematics, 27 (1998), pp. 533–557.

[67] J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics*, Geophysics, 74 (2009), pp. WCC1–WCC26.

[68] S. Wang, M. V. de Hoop, and J. Xia, *On 3d modeling of seismic wave propagation via a structured parallel multifrontal direct helmholtz solver*, Geophysical Prospecting, 59 (2011), pp. 857–873.

[69] S. Xu, D. Wang, F. Chen, and G. Lambaré, *Full waveform inversion for reflected seismic data*, Journal of Seismic Exploration, 22 (2013).

[70] Y. Yang, B. Engquist, J. Sun, and B. F. Hamfeldt, *Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion*, Geophysics, 83 (2018), pp. R43–R62.



RESEARCH CENTRE
BORDEAUX – SUD-OUEST

200 avenue de la Vieille Tour
33405 Talence Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399