Migration-based Traveltime Waveform Inversion of 2D Simple Structures: the Synclay Model

Guy Chavent, François Clément, Susana Gómez

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Migration-based Traveltime Waveform Inversion of 2D Simple Structures: the Synclay Model

Guy Chavent, François Clément, Susana Gómez

N ° 3502

Octobre 1998

THÈME 4

ISSN 0249-6399

INRIA
ROCQUENCOURT

Migration-based Traveltime Waveform Inversion of 2D Simple Structures: the Synclay Model

Guy Chavent*†, François Clément†, Susana Gómez‡

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Estime

Rapport de recherche n ° 3502 — Octobre 1998 — 28 pages

Abstract: A reformulation of the classical least-squares waveform inversion problem is proposed to retrieve the background velocity from surface seismic reflection data. To eliminate the problem of local minima, a change of unknown is performed: the depth reflectivity is replaced by a reflectivity in the time domain through a migration step. The increase in modeling complexity is the price for the enlargement of the attraction domain of the global minimum, which allows for the use of a local optimization method. A numerical implementation with a finite difference acoustic simulator is detailed. Inversion of a non-layered 2D slowness model is performed.

Key-words: inverse problem, time formulation, acoustic waves, finite difference, optimization, background velocity inversion, reflectivity inversion

(Résumé : tsvp)

This research was carried out as part of the Seismic Inversion, Geophysical Modeling and Applications Consortium. The authors hereby acknowledge the support provided by the sponsors of the SIGMA Consortium.

* CEREMADE, Université de Paris-Dauphine, pl. du Maréchal de Lattre de Tassigny, F-75775 Paris Cédex 16.
† Projet Estime
‡ IIMAS, National University of Mexico, Apdo. Postal 20-726, Mexico D.F.

Une formulation en temps de parcours par migration pour l'inversion par forme d'onde de structures 2D simples : le modèle Synclay

Résumé : Une reformulation du problème classique des moindres carrés d'inversion par forme d'ondes est proposée pour retrouver le modèle de vitesse à partir de données de surface de sismique-réflexion. Pour éliminer le problème de minima locaux, un changement d'inconnue de réflectivité est réalisé : la réflectivité en profondeur est remplacée par une réflectivité en temps par une étape de migration. L'augmentation de la complexité de la modélisation est le prix pour un élargissement du domaine d'attraction du minimum global, ce qui permet l'utilisation d'une méthode d'optimisation locale. Une implémentation avec un simulateur acoustique par différences finies est détaillée. L'inversion d'un modèle de lenteur 2D non stratifié est réalisée.

Mots-clé : problème inverse, formulation en temps, ondes acoustiques, différences finies, optimisation, inversion du modèle de vitesse, inversion de la réflectivité

1 Introduction

Waveform inversion of reflection seismic data has been impaired by the extremely poor behavior of the least-squares data misfit with respect to the low-frequency components of the slowness model: saturation effects and local minima have prevented local optimization algorithm to get close to the global minimum.

Various approaches have been tried to overcome this difficulty: frequency continuation [2], frequency and time continuation [10], both limited by the low low-frequency content of seismic data; global optimization [12, 15], limited by the large number of iterations required to approach the global minimum. Recently, two approaches have been proposed, which seem to have the potential of solving the problem: the Differential Semblance Optimization (DSO) [16] and the Migration-Based TravelTime (MBTT) formulation [6, 3]. The MBTT formulation has been successfully applied to the inversion of real VHR data using a simple Born+Ray modeling [14]. We shall investigate in this paper the capabilities of the MBTT formulation when a more complex finite difference acoustic forward modeling is used.

We recall first the principle of the MBTT formulation. The first observation is that, for 1D problems, the optimizability of the least-squares data misfit can be restored by searching for rock properties function of traveltime rather than depth [1]. The MBTT approach is based on the same philosophy, with two difficulties specific to 2D (or 3D) problems: there is no uniquely defined depth-to-traveltime transformation, and the background slowness, which plays no role in 1D propagation, enters the picture. The idea is then to minimize the same least-squares data misfit, but with respect to different optimization variables: instead of minimizing with respect to the slowness model ν, as one would do in the classical least-squares waveform inversion, one minimizes with respect to a "propagator" unknown p and a "time reflectivity" unknown s.

The propagator p defines the smooth slowness background $\nu_s(p)$, which is responsible for the kinematics effects. The time reflectivity s allows to store, in the time domain, the current information on the high frequency component ν_r of the slowness, which is responsible for the reflectivity effects. The new unknown s has exactly the same structure and dimension as the data d: if d is made of NS shotgathers $d_1, \dots, d_{\mathsf{NS}}$, so is s made of NS "shotgathers" $s_1, \dots, s_{\mathsf{NS}}$. The reconstruction of the depth section ν_r from the time section s requires the knowledge of the current propagator p; it is performed by prestack depth migrating each shotgather s_n with the background $\nu_s(p)$, and stacking the migrated shotgathers. This results in a depth reflectivity vector r given by

$$r = \sum_{n=1}^{\mathsf{NS}} \boldsymbol{\mathcal{M}}_n(p) s_n \tag{1}$$

where $\boldsymbol{\mathcal{M}}_n(p)$ is a prestack depth migration operator for shot n with the background slowness $\nu_s(p)$.

If we denote by $\nu_r(r)$ the rough slowness corresponding to the depth reflectivity r, we can finally associate to any current value of the propagator p and of the time reflectivity s a slowness distribution ν by

$$\nu = \nu_s(p) + \nu_r(r) \tag{2}$$

with r computed by (1).

Finally, the Migration-Based TravelTime reformulation of the least-squares waveform inversion consists in performing in this latter the $\nu \mapsto (p, s)$ change of unknown defined by (1) and (2). This includes replacing the search for the depth reflectivity r by the search for a time reflectivity s, which is the 2D generalization, based on migration, of the 1D depth-to-traveltime transformation; and explains the name given to this reformulation.

The paper is organized as follows. We recall shortly and illustrate numerically in the first section the limitations of the classical least-squares approach. In the second section, we describe the

implementation of the MBTT approach for the inversion of seismic data using a 2D finite difference acoustic simulator. In particular, we describe the smooth/rough parameterization of the slowness and the prestack depth migration operator which are the basis of the MBTT reformulation, and we discuss the implementation and computational cost of MBTT. We present the optimization strategy in the third section. In particular, we take advantage of a multiscale representation of the propagator. And in the forth and last section, we actually minimize the least-squares waveform misfit by a local gradient method. This allows to invert synthetic data for the full slowness (both background and high frequency component) in situations where the classical least-squares formulation fails. The slowness reconstructed by the MBTT inversion is then used as initial guess in a classical least-squares minimization, which provides the final touch to the inverted slowness section. Finally, we analyze numerically the influence of the initial guess and of the number of shotgathers on the quality of the inverted slowness model. We have illustrated all the sections with numerical results obtained with the synthetic Synclay model described in Appendix A. The other appendices are devoted to a full description of the finite difference acoustic simulator and the migration, the smooth/rough decomposition, and the computation of the gradient of the least-squares misfit after MBTT reformulation.

2 Classical least-squares inversion and its limitations

Let $d = (d_1, \dots, d_{\mathsf{NS}})$ a collection of shotgathers to be inverted, $\nu(x,z)$ the current slowness map and $c = (c_1, \dots, c_{\mathsf{NS}})$ the corresponding collection of synthetic shotgathers. The least-squares waveform data misfit J associated to ν is

$$J = \sum_{n=1}^{\mathsf{NS}} J_n = \sum_{n=1}^{\mathsf{NS}} \frac{1}{2} \|d_n - c_n\|^2_{WG}, \tag{3}$$

where $\|.\|_{WG}$ is a weighted Euclidean norm on shotgathers (it includes a summation on all receivers and all time samples). The weight WG accounts for the muting of the direct arrivals and for the amplitude correction for the geometrical spreading.

The Classical Least-Squares (CLS) formulation of waveform inversion is:

find $\hat{\nu}$ which minimizes J given by (3) where the c_n are computed with the slowness ν.

It is well-known that this formulation, though very satisfactory in principle, suffers from a major flaw: J has numerous local minima with respect to the low-frequency components of the slowness ν.

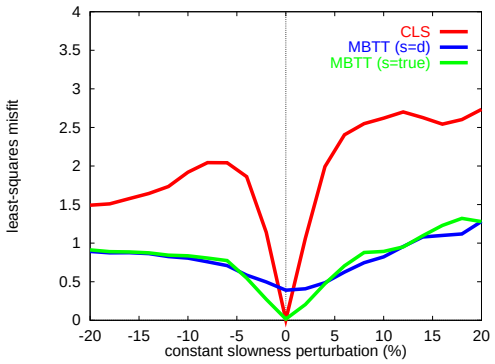

Figure 1: Scans of the least-squares misfit for constant perturbations of the slowness (in percentage of its mean value). The least-squares misfit is normalized by the value $\frac{1}{2}\|d\|^2_{WG}$.

This difficulty is illustrated on Figure 1 where the CLS curve (in red) shows that local minima are present for constant slowness perturbations as small as 15%. As a consequence, minimization

of J with a local gradient method is unfeasible, unless the initial guess of slowness is within a few percents of the true solution, which is unrealistic for any practical purpose. This can be seen on Figure 9 where the slowness model obtained by the CLS approach with a quasi-Newton optimizer does not depart from its constant initial value!

3 Migration-based traveltime with finite difference acoustic modeling

We consider in this paper a forward modeling by a 2D finite difference acoustic simulator (described in Appendix B). We suppose that the density is smooth and known, and we want to retrieve the slowness model $\nu(x,z)$ from the collection of shotgathers $d = (d_1, \ldots, d_{\mathsf{NS}})$.

In order to perform the MBTT change of unknown, one has first to separate in ν the propagation and reflection effects. It is well-known that these effects are governed respectively by the low and high spatial frequency components of ν, with the intermediate frequencies having little influence on the data ("frequency gap" reported in [9]).

Let us denote by V the space of bilinearly interpolated functions on the finite difference grid of the acoustic simulator, where the slowness $\nu(x,z)$ lives. We perform first a smooth/rough decomposition of ν by choosing two subspaces V_s and V_r of V such that: (i) V_s is made of slowly varying ("smooth") functions, (ii) V_r is made of rapidly oscillating ("rough") functions and (iii) V is the direct sum of V_s and V_r.

Our implementation of this is based on a multiscale decomposition of the space V. If the grid for V is the N-th dyadic refinement of a coarse initial grid made of a few elements only, then one can choose for V_s the space of bilinearly interpolated functions on the grid obtained by N_s divisions of the coarse grid ($0 \leq N_s < N$), and for V_r a complement to V_s in V. We refer to Appendix C for a precise description of the multiscale decomposition of V, and of the basis functions of V_s and V_r.

The number N_s of scales of the slowness space V which are used to build the space V_s of smooth slownesses is determined by numerical experimentation where we require that:

- a synthetic generated with a smooth slowness ν_s in V_s does not contain sizeable reflections,
- adding a rough slowness ν_r in V_r to a smooth slowness ν_s in V_s does not alter significantly the traveltimes.

A way to ensure this last property for medium of limited complexity is to choose for V_r the orthogonal complement to V_s in V, rather than the natural (oblique) complement associated to the multiscale decomposition of V.

We have illustrated in Figure 2 the smooth/rough decomposition on the Synclay slowness model of Figure 11. Notice that the sum of the smooth and rough components reconstructs exactly the full initial slowness.

Once N_s has been chosen, *i.e.*, the smooth/rough decomposition of V has been chosen, we define the propagator unknown p and the depth reflectivity unknown r by

p = coefficients of ν_s on the basis of V_s,
r = coefficients of ν_r on the basis of V_r.

So to any propagator p and depth reflectivity r, we associate a slowness ν by

$$\nu = \nu_s(p) + \nu_r(r). \tag{4}$$

The least-squares misfit J can now be considered as a function of the two variables p and r.

Notice that, thanks to the above mentioned "frequency gap", the choice of N_s is not so critical. Notice also that, by construction, the space V_s itself has a multiscale structure, as it contains scales $0, 1, \ldots, N_s$. This will prove extremely useful for the determination of the propagator p (*i.e.*, the background slowness $\nu_s(p)$): we shall use a multiscale optimization with respect to p,

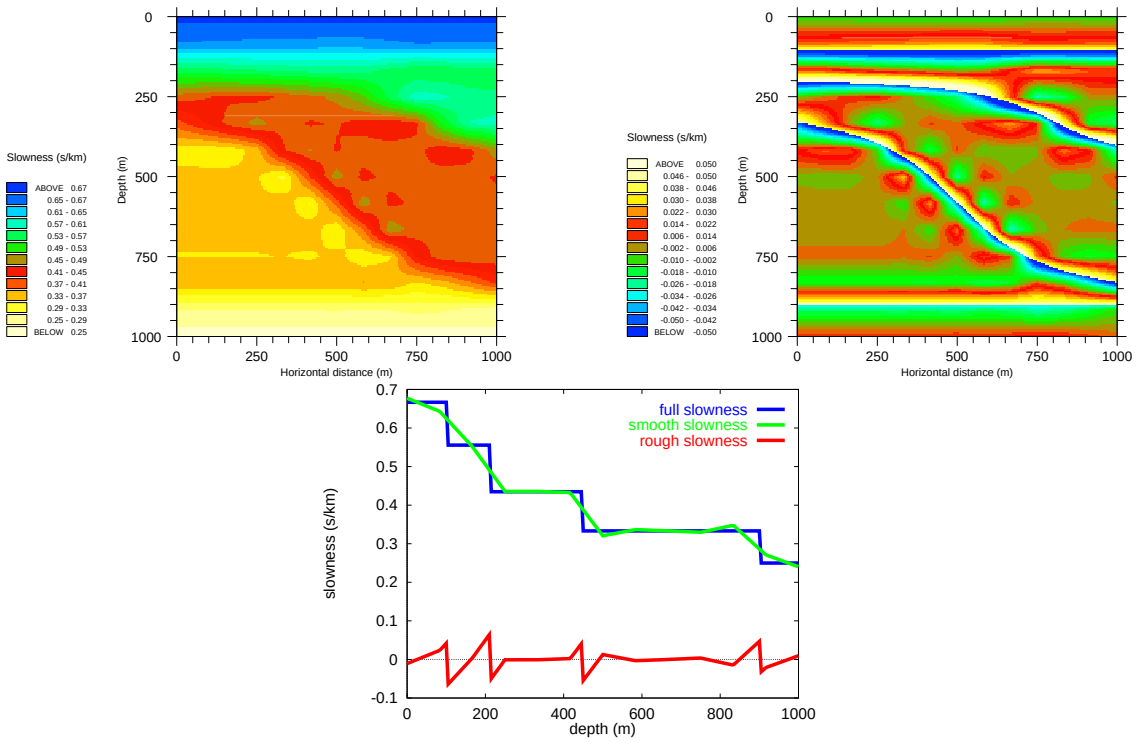

Figure 2: Smooth/rough decomposition of the Synclay slowness model. Smooth slowness (top left) and rough slowness (top right). Profiles at horizontal distance 350 m (bottom).

searching first for the general trend of the slowness (at scale 0), and then allowing more and more details to enter the picture (introducing the scales $1, \dots, N_s$).

As we recalled in the introduction, the principle of the Migration-Based TravelTime approach consists in parameterizing the depth reflectivity r by a time reflectivity s, from which r can be deduced by migration. It is hence important to use there a true amplitude migration, as this will ensure that the data themselves are a good initial guess for the time reflectivity s.

Let $\boldsymbol{\mathcal{F}}_n$ be the forward map which, to any p, r associates the n-th synthetic shotgather c_n computed by the finite difference code from the slowness given by (4),

$$c_n = \boldsymbol{\mathcal{F}}_n(p, r) \qquad n = 1, 2, \dots, \mathsf{NS}.$$

It is now well-known [11, 17, 18] that

$$\nabla_r J_n = -\left(\frac{\partial \boldsymbol{\mathcal{F}}_n}{\partial r}(p,0)\right)^T (d_n - \underbrace{\boldsymbol{\mathcal{F}}_n(p,0)}_{\simeq 0})$$

is a prestack two-way reverse-time depth-migrated section for the n-th shotgather, which in general does not preserves the amplitudes of deep events. So our choice for the true amplitude migration operator $\boldsymbol{\mathcal{M}}(p)$ in charge of transforming the time reflectivity $s = (s_1, \dots, s_{\mathsf{NS}})$ into the depth reflectivities $r_1, \dots, r_{\mathsf{NS}}$ and stacking the result is

$$\boldsymbol{\mathcal{M}}(p) = WM \left(\frac{\partial \boldsymbol{\mathcal{F}}_1}{\partial r}(p,0)^T \dots \frac{\partial \boldsymbol{\mathcal{F}}_{\mathsf{NS}}}{\partial r}(p,0)^T \right), \tag{5}$$

where WM is a diagonal weight matrix to restore the amplitudes of the migrated events. We refer to Appendices B and C for the calculation of the action of $\frac{\partial \boldsymbol{\mathcal{F}}_n}{\partial r}(p,0)^T$ on any time section.

The MBTT change of reflectivity unknown writes now

$$r = \mathcal{M}(p)s = WM \sum_{n=1}^{\mathsf{NS}} \frac{\partial \mathcal{F}_n}{\partial r}(p,0)^T s_n, \tag{6}$$

so that the synthetic c associated now to the propagator p and to the time reflectivity s is given by

$$c = \mathcal{F}(p,r) = \mathcal{F}(p,\mathcal{M}(p)s). \tag{7}$$

We shall say that c is the **resimulation of s with the propagator p.**

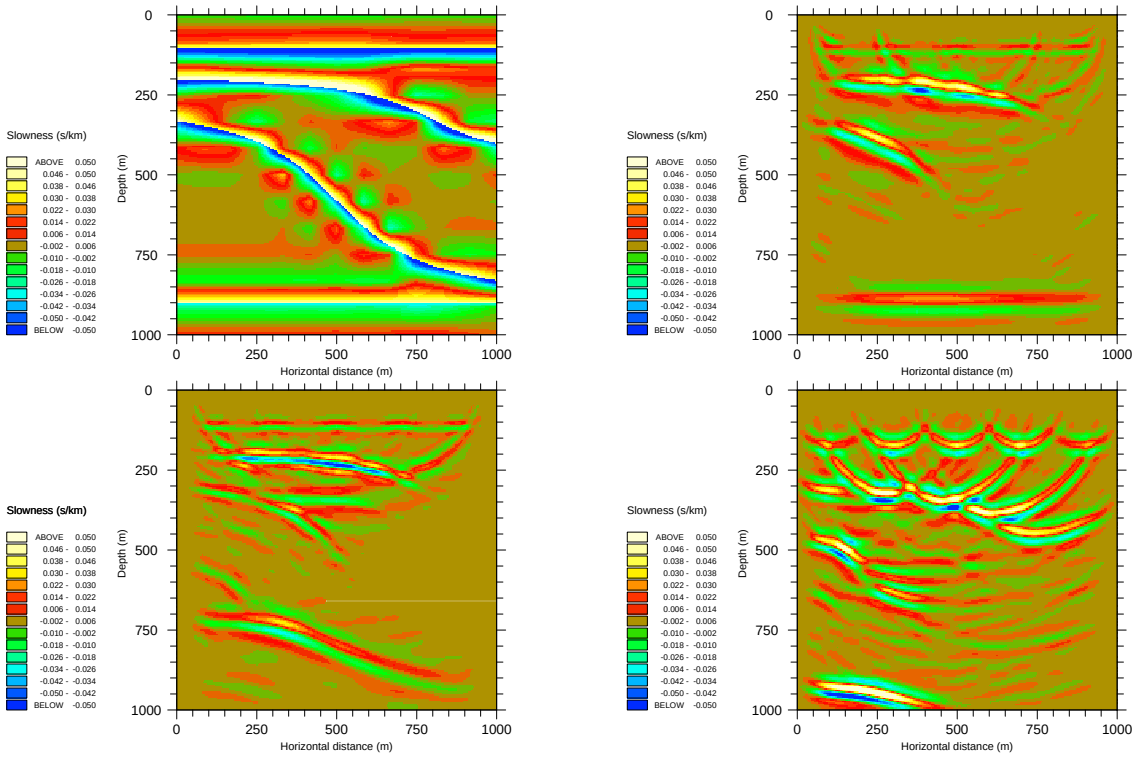

Figure 3: True rough slowness (top left). Quantitative migration of the data with the true propagator (top right), with the ramp propagator (bottom left) and with the constant propagator (bottom right).

We have illustrated in Figure 3 the quantitative migration of the Synclay data of Figure 11 ($s = d$) successively with the true Synclay propagator of Figure 2, with a ramp propagator (which has the true surface and bottom values), and with a constant propagator (which has the true mean value). We see that the stack over the individually prestack migrated shotgathers is more constructive with the true propagator (top right) than with the wrong ramp propagator (bottom left), which is itself more constructive than with the very wrong constant propagator (bottom right).

If we look now at the time sections of Figure 4 obtained by forward modeling the migrated sections of Figure 3 with the corresponding propagators, *i.e.*, obtained by resimulation of the Synclay data of Figure 11 with the true, ramp and constant propagators, we see that the better coherency of the migrated section for the true propagator (top right) produces a time section whose events have amplitudes closer to the data than those of the time section generated by the less coherent migrated section for the wrong ramp propagator (bottom left), themselves closer

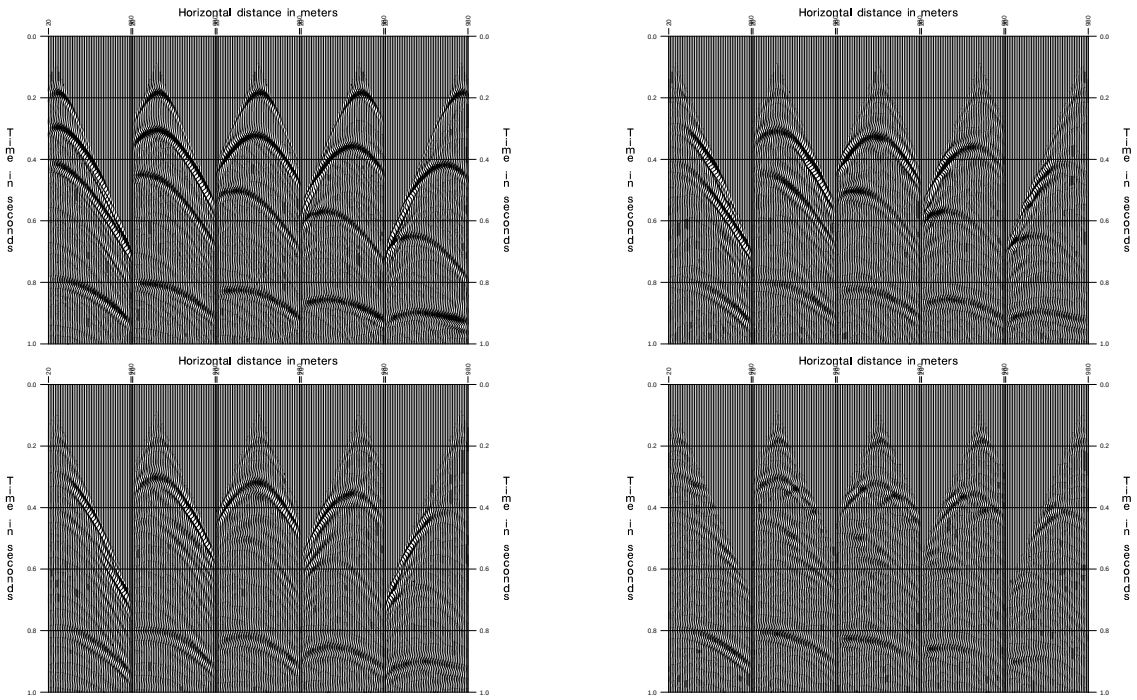

Figure 4: 5-shot data (top left). 5-shot resimulation of the data with the true propagator (top right), with the ramp propagator (bottom left) and with the constant propagator (bottom right). Respectively 38%, 27% and 11% of the data are explained.

than those of the time section generated by the lesser coherent migrated section for the very wrong constant propagator (bottom right). Hence the coherency of the migrated section is translated in term of amplitudes of the resimulated section in the time domain.

We also see that the MBTT resimulation forward map eliminates the phase shifts problem: the events in the resimulated sections of Figure 4 are at the same time location, though these sections are evaluated with quite different background propagators. As a consequence, one can expect that the domain of attraction of the global minimum for the least-squares data misfit (3) will be significantly enlarged by the MBTT change of unknown (Figure 1, blue curve).

The MBTT reformulation of the least-squares waveform inversion is then:

find $\hat{p}, \hat{s}$ which minimize J given by (3) where c is resimulated from p, s by (7).

The numerical results presented in the last section confirm the expected enlargement, as we have been able to retrieve, on synthetic data, the global minimum by a local gradient method starting from very poor initial guess of the propagator.

Another plus of the MBTT formulation results from our choice of an amplitude preserving migration for $\mathcal{M}(p)$: if p_{true} and r_{true} denote the propagator and depth reflectivity which have generated the data d, then $\mathcal{M}(p_{\text{true}})d$ will be close to r_{true} where the Earth is illuminated by the seismic experiment. This suggests that the data d themselves are in fact an approximation of the true time reflectivity s_{true}, if such time section exists!

The problem of the existence of a time reflectivity s_{true} such that the resimulation of s_{true} with p_{true} coïncides with d is mathematically solved when a linearized forward model is used [14]. But with the full acoustic forward modeling we use here, the previous theoretical result does not apply. So we have checked numerically, on the synthetic Synclay data d of Figure 11, the existence of an s_{true} whose resimulation with p_{true} was very close to the data d. The time reflectivity s_{true} is shown in Figure 5 where one sees that it is a deconvolved and de-multiplied version of the data,

as its migration (which involves a convolution with the source) is very close to the true depth reflectivity.

We have now completed the description of the MBTT change of unknown, up to the specifications of the weight WM in charge of restoring the "true amplitude" in the migration operator $\boldsymbol{\mathcal{M}}(p)$ defined in (5). A mathematically sound choice for WM would be an approximation to the inverse of the Hessian of J with respect to r [8, 5]. But the computation of the Hessian is computationally extremely expensive. Moreover, we do not need in fact the ultimate WM, as optimization with respect to s will correct for possible defaults in WM: if WM underestimates the amplitude of some migrated events (as it is clearly the case in Figure 3, top right, for the water bottom reflection), the optimization with respect to s will boost the amplitude of the corresponding time events of s (as it can be seen in Figure 5 for the water bottom reflection), in such a way that the migrated event returns to its correct amplitude. This is why we have chosen for WM in this experiment a very simple weight proportional to the identity matrix,

$$WM = \lambda \mathbf{Id}, \tag{8}$$

where, for each propagator p, the coefficient λ is determined by requiring that the corresponding least-squares misfit J is minimum.

We conclude this section by commenting the increase in complexity and computational intensity incurred by the MBTT reformulation.

First, the dimension of s (the same as d) is much larger than that of r. But this is not really a problem, as the objective function is nearly quadratical with respect to s (as with respect to r), so that conjugate gradient type algorithms can be used.

Second, the evaluation of the data misfit J, given one set of parameters p and s, requires the computation of a prestack depth migration of s, followed by a forward modeling. This increase in complexity not only multiplies by 5.4 the computation time for the evaluation of J, but also makes extremely cumbersome the determination of the adjoint equations for the computation of the analytical gradient of J with respect to p and s (to get an idea, look at Appendix D, where we desperately tried to make it look simple). Of course, the coding of those equations is still more delicate: a lot of debugging is needed before the analytical gradient given by the adjoint equations coïncides with the finite difference gradient obtained by perturbing slightly each parameter one at a time. But once this is done, the computational cost for the evaluation of the gradient of J with respect to p and s is also 5.2 times larger than the cost of the gradient with respect to ν.

4 Optimization strategy

In the MBTT approach, one has to minimize the least-squares misfit J defined by (3) with respect to the two unknowns p (propagator) and s (time reflectivity), with the propagator p being itself decomposed on a multiscale basis. So there are many possible ways to intertwine the optimizations with respect to p at various scales, and with respect to s. After a systematic experimentation, we have retained the strategy which has revealed the most effective: **the multiscale alternate strategy** (MA-MBTT), where at each scale one optimizes first with respect to p and then with respect to s (at scale 0, s is initialized by d), the computational effort (number of minimization iterations) being decreased for p and increased for s as the scale goes from 0 to $N_s = 2$. An additional CLS minimization with respect to ν has also been performed using as initial value the slowness $\nu = \nu_s(p) + \nu_r(r)$ with $r = \boldsymbol{\mathcal{M}}(p)s$ reconstructed from the optimal MBTT values of p and s.

In order to allow for comparison, we have also implemented the **classical least-squares** (CLS) strategy, which consists in minimizing the misfit J directly with respect to the slowness model ν.

All numerical experimentations have been made with a fixed number of iteration with respect to each variable (index of p refers to the scale at which the optimization is performed):

MA-MBTT+CLS $30/p_0$, $5/s$, $20/p_1$, $10/s$, $10/p_2$, $15/s$, $40/\nu$,

CLS $600/\nu$

The number of iterations for the CLS strategy has been chosen in such a way that the computation time is roughly the same for both strategies.

All the above minimizations have been performed using the `m1qn3` routine developed at INRIA. This minimizer solves unconstrained problems by a variable-storage quasi-Newton method. At each iteration, a descent direction is determined from an approximation of the inverse of the Hessian, then a step length is determined along that direction by a line-search procedure. The approximation of the inverse of the Hessian is obtained through a (small) fixed number of iterations of the inverse BFGS formulas. More details can be found in [7, 13].

5 Inversion results

We have inverted the synthetic Synclay data set described in Appendix A, first with 5 shots (Figure 11), and then with 41 shots.

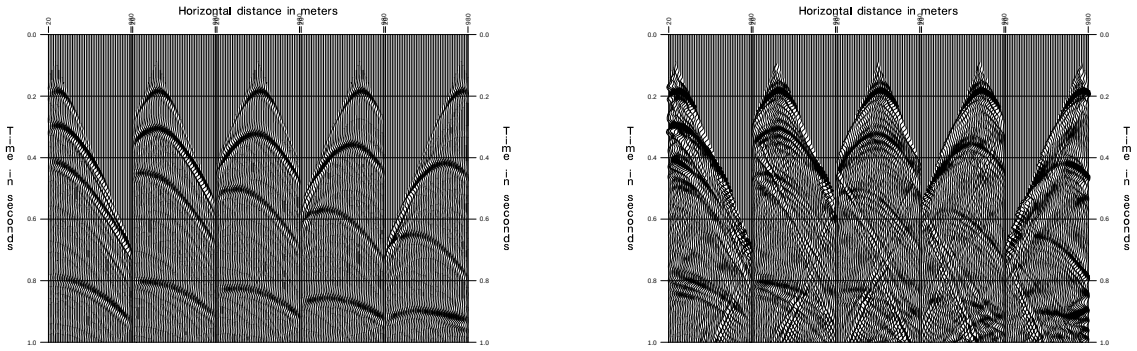

Figure 5: 5-shot data (left). 5-shot true time reflectivity s_{true} (right).

We have first minimized J with respect to the time reflectivity s for $p = p_{\text{true}}$ given by Figure 2 (top left). This resulted in the time reflectivity of Figure 5, whose resimulation gave a very good fit to the data as 93% of the data are explained. Hence it can be considered as the true time reflectivity s_{true}. This shows that the MBTT change of variable is not under-parameterizing, and gives a numerical justification to the MBTT change of unknown when a finite difference acoustic modeling is used.

Then, we have inverted the 5-shot Synclay data set with the MA-MBTT+CLS strategy, with the linear ramp as initial guess. Figure 6 summarizes the results of this inversion. The decrease of the objective function is to be seen at the top left, the optimal propagator p at the top right and the optimal time reflectivity s at the middle left. Then, at the middle right, we depict the rough part ν_r estimated by migration, with the previous background and time reflectivity. One can see that the reflectors are correctly located, even the deepest one. Adding the smooth and rough parts provides then the initial guess (bottom left) for the final touch CLS optimization, which results in the slowness model at the bottom right. This slowness model explains 91% of the data.

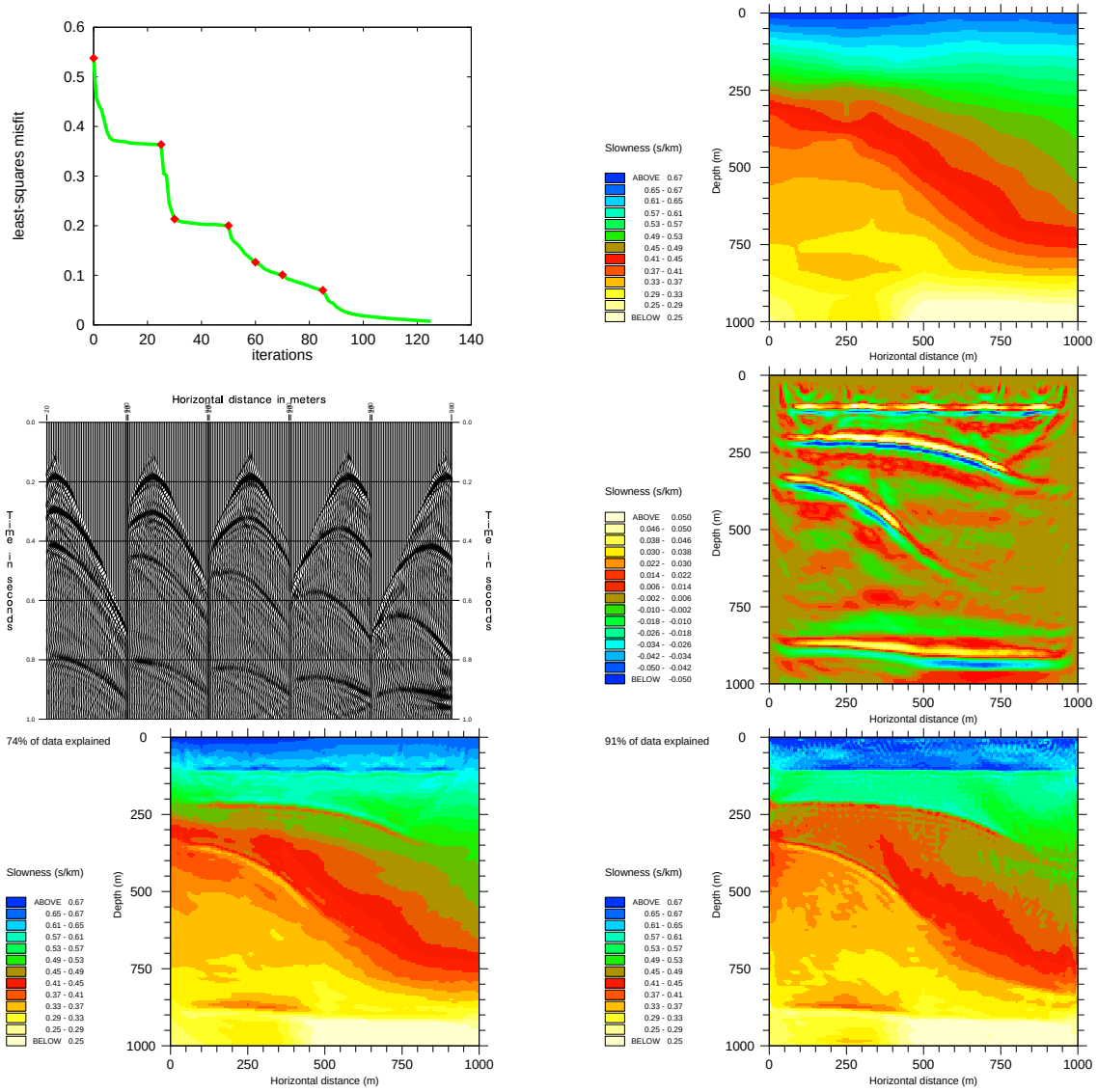

Figure 6: Inversion of the 5-shot Synclay data set using the MA-MBTT+CLS strategy starting from the ramp slowness. Decrease of the least-squares misfit (top left). Propagator after the last minimization with respect to p (top right) and 5-shot time reflectivity after the last minimization with respect to s (middle left). Corresponding depth reflectivity (middle right) and reconstructed slowness (bottom left). Slowness after minimization with respect to ν starting from the reconstructed slowness (bottom right).

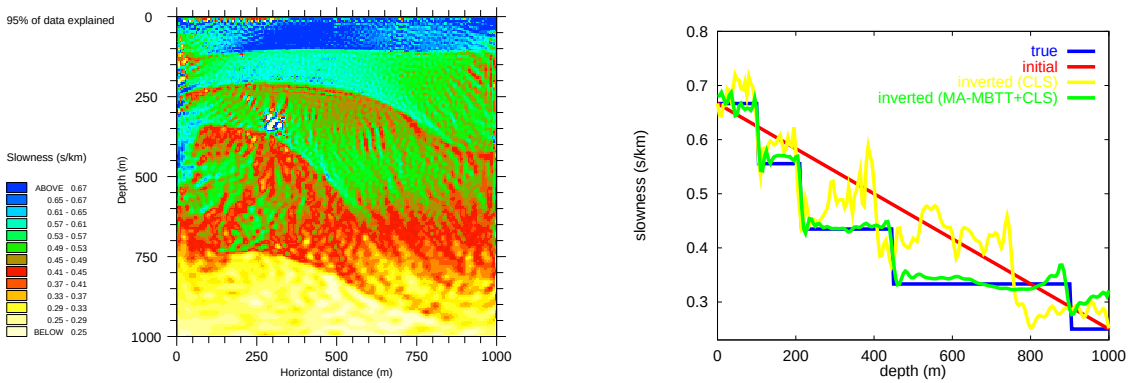

Figure 7: Inversion of the 5-shot Synclay data set starting from the ramp slowness. Slowness after minimization with respect to ν using the CLS strategy (left). Slowness profiles at horizontal distance 350 m after inversion (right).

For comparison, we show on Figure 7 (left) the result of the CLS strategy for the same problem: as the initial ramp contains already a large part of the low frequency of the true slowness model, the quasi-Newton algorithm has been able to get close to the global minimum of the objective function, and the slowness of Figure 7 (left) explains 95% of the data, even slightly more than the MA-MBTT+CLS strategy.

Figure 8 summarizes the results of the inversion of the same 5-shot Synclay data with the same initial ramp (top left) by the CLS technique (bottom left) and the MA-MBTT+CLS strategy (bottom right) and Table 1 summarizes the optimization reports: the decrease of the misfit, the number of iterations, the number of evaluations of the function and its gradient, and the computational time.

Figure 9 shows the influence of a poor initial guess (the constant slowness of 0.4 s/km) on the inversion of the 5-shot Synclay data set: the MA-MBTT+CLS strategy (bottom right) is still able to restore a part of the low frequencies of the slowness, and explains 89% of the data. In comparison, the CLS approach (bottom left) is unable to move the slowness away from its initial value, and explains only 61% of the data.

Finally, we investigate in Figure 10 whether an increase in the shot number from 5 to 41 helps in retrieving a correct slowness model. The slowness model obtained by MA-MBTT+CLS (bottom right) is much better than the one obtained in Figure 9 with only 5 shots, it explains 96% of the data, the best performance of our tests. On the other end, the increase in information seems to hurt the CLS approach, which is only able to explain 50% of the data, compared to 61% in the case of 5 shots!

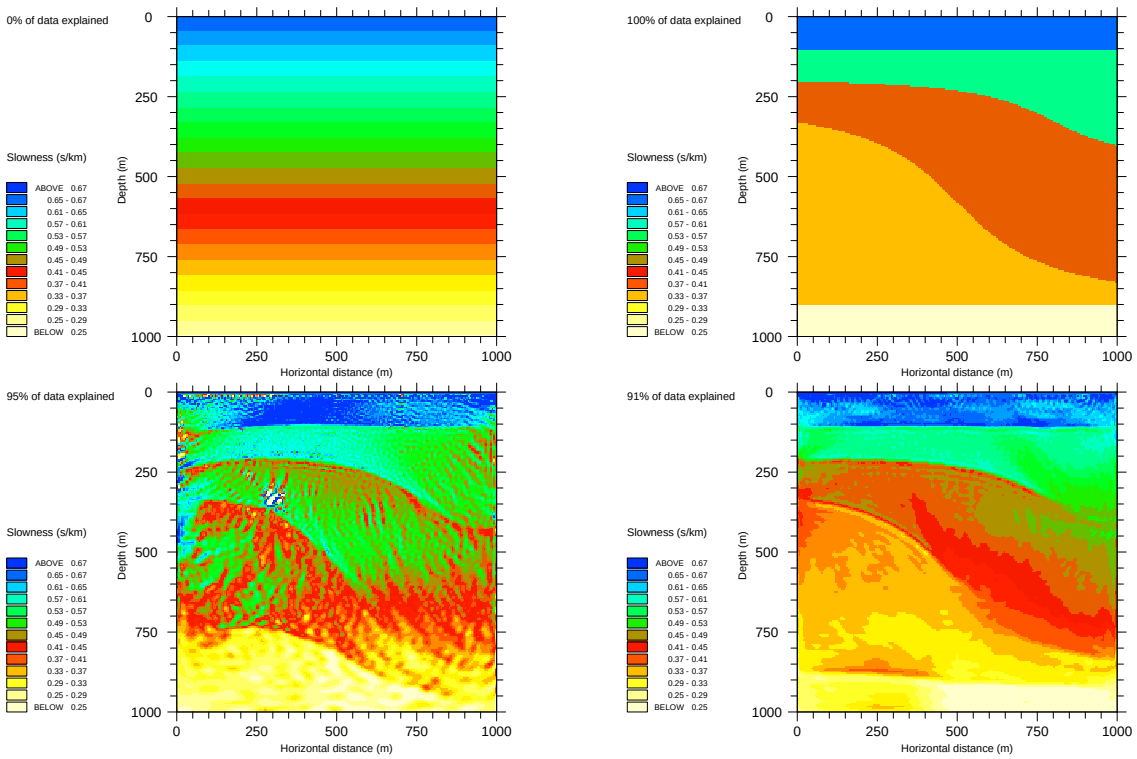

Figure 8: Inversion of the 5-shot Synclay data set using CLS and MA-MBTT+CLS strategies starting from the ramp slowness. Top: initial slowness (left), true slowness (right). Bottom: Inverted slowness using CLS (left) and using MA-MBTT+CLS (right).

strategy	normalized misfit	iterations	evaluations	cpu (h)
CLS	$1 \searrow 0.0030$	0+600	0+646	0+19
MA-MBTT+CLS	$0.54 \searrow 0.0073$	85+40	105+42	19+1

Table 1: Optimization reports for the inversion of the 5-shot Synclay data set using CLS and MA-MBTT+CLS strategies starting from the ramp slowness.

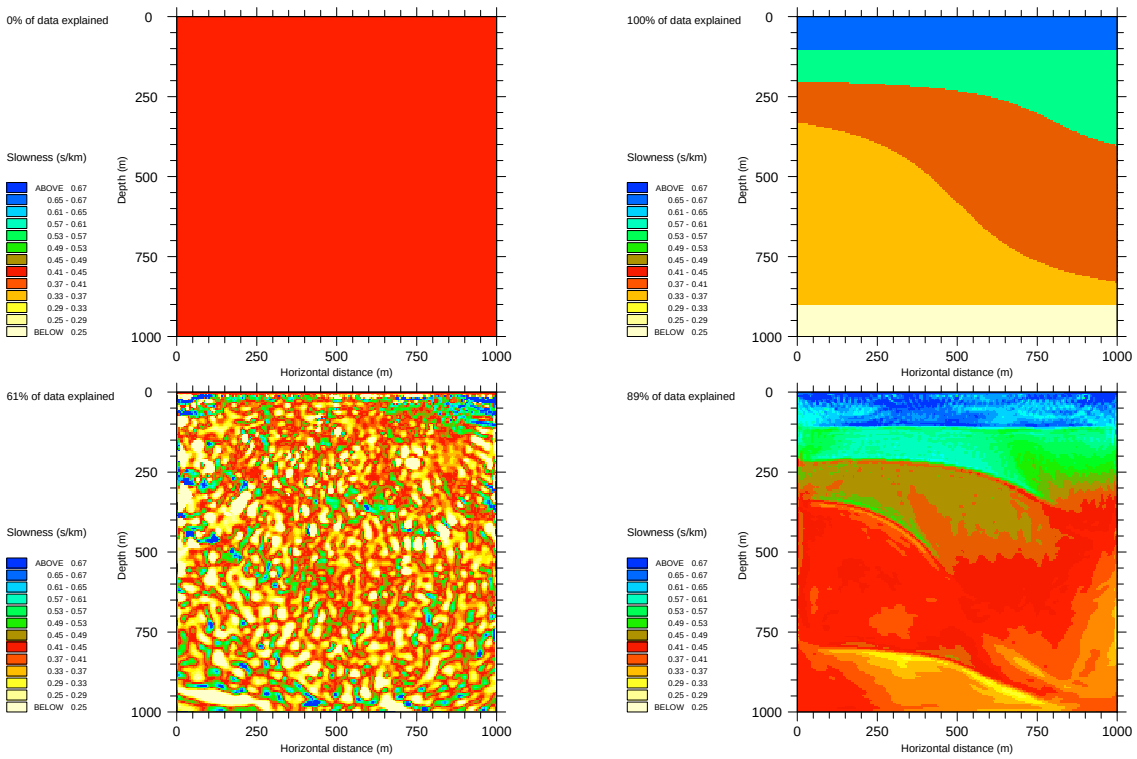

Figure 9: Inversion of the 5-shot Synclay data set using CLS and MA-MBTT+CLS strategies starting from the constant slowness. Top: initial slowness (left), true slowness (right). Bottom: Inverted slowness using CLS (left) and using MA-MBTT+CLS (right).

strategy	normalized misfit	iterations	evaluations	cpu (h)
CLS	$1 \searrow 0.152$	0+500	0+536	0+16
MA-MBTT+CLS	$0.79 \searrow 0.012$	90+40	108+42	19+1

Table 2: Optimization reports for the inversion of the 5-shot Synclay data set using CLS and MA-MBTT+CLS strategies starting from the constant slowness.

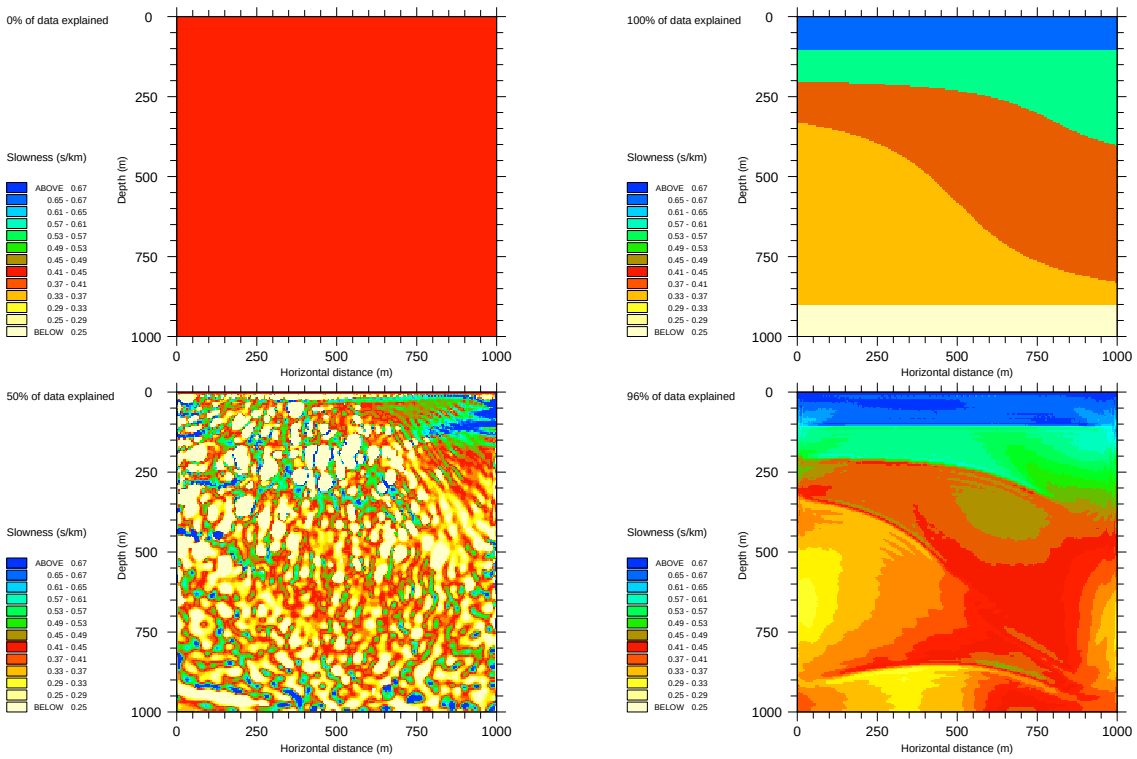

Figure 10: Inversion of the 41-shot Synclay data set using CLS and MA-MBTT+CLS strategies starting from the constant slowness. Top: initial slowness (left), true slowness (right). Bottom: Inverted slowness using CLS (left) and using MA-MBTT+CLS (right).

strategy	normalized misfit	iterations	evaluations	cpu (h)
CLS	$1 \searrow 0.252$	0+600	0+649	0+160
MA-MBTT+CLS	$0.91 \searrow 0.0020$	90+40	110+44	150+10

Table 3: Optimization reports for the inversion of the 41-shot Synclay data set using CLS and MA-MBTT+CLS strategies starting from the constant slowness.

6 Conclusion

The Migration-Based TravelTime (MBTT) approach to full waveform inversion eliminates the phase shifts on the synthetics, which are responsible for the presence of local minima for the classical least-squares (CLS) data misfit function. Relying on a change of reflectivity unknown trough a quantitative migration step from the CLS formulation, it allows the computation of a gradient with a reasonable overcost. This opens the way to minimization by local methods.

The MBTT change of reflectivity unknown is justified numerically. We have successfully inverted a synthetic model which presents lateral slowness variations: when using a very poor constant initial guess for the slowness, the minimization of the MBTT misfit restores correctly the main trends of the true slowness model in the illuminated areas, while the CLS version cannot depart from the initial constant.

This MBTT approach based on a 2D finite difference acoustic simulator seems very promising towards the solution of the background slowness inversion for real applications, in particular for CMP gathers containing multiple events.

References

[1] A. Bamberger, G. Chavent, and P. Lailly. About the stability of the inverse problem in 1D wave equations. Application to the interpretation of seismic profiles. *Appl. Math. Optim.*, 5:1–47, 1979.

[2] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent. Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, 1995.

[3] G. Chavent, F. Clément, and S. Gómez. Automatic determination of velocities via migration-based traveltime waveform inversion: a synthetic data example. In *Proc. of the 64th S.E.G. Annual Meeting*, pages 1179–1182. Soc. of Expl. Geophys., 1994. Expanded abstracts.

[4] G. Chavent and C. A. Jacewitz. Automatic determination of background velocities by multiple migration fitting. *Geophysics*, 60(2), 1995.

[5] G. Chavent and R.-E. Plessix. An optimal true-amplitude least-squares prestack depth migration operator. *Geophysics*, To appear.

[6] F. Clément. A migration-based travel-time formulation for the inversion of 2D seismic reflection data. In G. Cohen, L. Halpern, and P. Joly, editors, *Proc. of the 1st Internat. Conf. on Mathematical and Numerical Aspects of Wave Propagation Phenomena*, pages 455–461. SIAM, 1991.

[7] J.-C. Gilbert and C. Lemaréchal. Some numerical experiments with variable storage quasi-newton algorithms. *Math. Programming*, 45(3):407–435, 1989.

[8] S. H. Gray. True-amplitude seismic migration: a comparison of three approaches. *Geophysics*, 62(3):929–936, 1997.

[9] M. Jannane, W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, D. Trezeguet, and M. Xie. Wavelength of earth structures that can be resolved from seismic reflection data. *Geophysics*, 54(7):906–910, 1989. Short Note.

[10] P. Kolb, F. Collino, and P. Lailly. Prestack inversion of a 1-D medium. *Proc. of the IEEE*, 74(3), 1986. Special Issue on Seismic Inversion.

[11] P. Lailly. The seismic inverse problem as a sequence of before stack migrations. In *Proc. of the Conf. on Inverse Scattering, Theory and Applications*. SIAM, 1983.

[12] E. Landa, W. Beydoun, and A. Tarantola. Reference velocity model estimation from prestack waveforms: Coherency optimisation by simulated annealing. *Geophysics*, 54(8):984–990, 1989.

[13] D. C. Liu and J. Nocedal. On the limited memory BFGS methods for large scale optimization. *Math. Programming*, 45(3):503–528, 1989.

[14] R.-E. Plessix, Y.-H. De Roeck, and G. Chavent. Waveform inversion of reflection seismic data for kinematic parameters by local optimization. *SIAM J. on Scientific Computing*, To appear.

[15] M. K. Sen and P. L. Stoffa. *Global optimization methods in geophysical inversion.* Elsevier, 1995.

[16] W. W. Symes and J. J. Carazzone. Velocity inversion by differential semblance optimization. *Geophysics*, 56(5), 1991.

[17] A. Tarantola. Inversion of seismic reflection data with acoustic approximation. *Geophysics*, 49:1259–1266, 1984.

[18] A. Tarantola. Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015, 1984.

A The Synclay data set

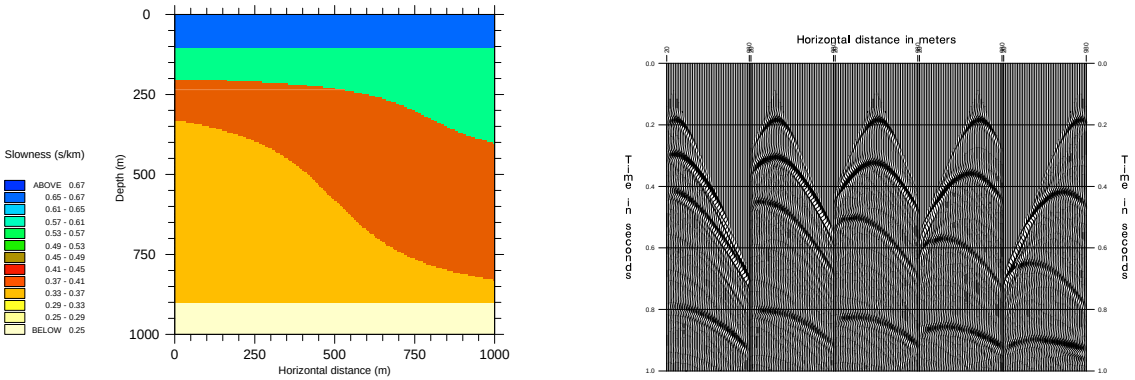

Figure 11: Synclay model: slowness (left), 5-shot data set (right).

The Synclay model is a synthetic 2D acoustic model. It is composed of five non horizontal layers leaning on a synclinal fold, with a shallow layer made of water (Figure 11, left). The velocity distribution and the density distribution are defined on a rectangular grid of 201×201 points with a resolution of 5 m. The velocity takes the values 1500/1800/2300/3000/4000 (in m/s) and the density is constant of value 1000 kg/m^3.

The sources are Gaussian wavelets located at a depth of 10 m and equally distributed: with a spacing of 200 m for the 5-shot set (Figure 11, right), and with a spacing of 20 m for the 41-shot set. The downward propagating signal has a peak frequency of 22 Hz and a cut frequency of 70 Hz. The 49 receivers are located at the same depth of 10 m, with a spacing of 20 m. They are fixed for all the shots. The recording time is of 1 s and is sampled at 0.004 s.

B Direct and adjoint acoustic wave equations

We describe, as concise as possible, the fully discretized equations we use to compute the solution of the direct and adjoint acoustic wave equations. We use similar notations to those used in [4] (appendix B). The main difference is the use of second order absorbing boundary conditions.

B.1 Discretization and shorthand notations

B.1.1 Spatial discretization

The computational domain Ω is covered with a regular rectangular grid of $\mathsf{NX} \times \mathsf{NZ}$ cells of size $\Delta x \times \Delta z$. We denote by Ω_h the set of all grid nodes, and by Ω_h^0 the subset of nodes that are not on the upper boundary. We denote by Γ_h the set of all boundary nodes, and by Γ_h^0 the subset of boundary nodes that are not on the upper boundary. The number of nodes in each set is

$$|\Omega_h| = (\mathsf{NX}+1)(\mathsf{NZ}+1), \qquad |\Omega_h^0| = (\mathsf{NX}+1)\mathsf{NZ},$$
$$|\Gamma_h| = 2\mathsf{NX}+2\mathsf{NZ}, \qquad |\Gamma_h^0| = \mathsf{NX}+2\mathsf{NZ}-1.$$

Wavefields will be sought at nodes $M \in \Omega_h^0$ only, with the convention that they take the value zero at all nodes on the upper boundary. Auxiliary fields will be sought at nodes $M \in \Gamma_h^0$ only, with the convention that they take value zero at all nodes on the upper boundary and at all nodes in the interior. Parameter fields will be defined at all grid nodes $M \in \Omega_h$.

For any wavefields $\mathbf{u} = (u_M, M \in \Omega_h^0)$ and $\mathbf{w} = (w_M, M \in \Omega_h^0)$, for any auxiliary fields $\boldsymbol{\varphi} = (\varphi_M, M \in \Gamma_h^0)$ and $\boldsymbol{\psi} = (\psi_M, M \in \Gamma_h^0)$, and for any parameter fields $\boldsymbol{\alpha} = (\alpha_M, M \in \Omega_h)$ and $\boldsymbol{\beta} = (\beta_M, M \in \Omega_h)$, we define

1. scalar products $\langle .,. \rangle_{\Omega_h^0}$, $\langle .,. \rangle_{\Omega_h}$, $\langle .,. \rangle_{\Gamma_h^0}$, $\langle .,. \rangle_{\Gamma_h}$ and $\langle .,. \rangle_{\Omega_h^0 \times \Gamma_h^0}$

$$\langle \mathbf{u}, \mathbf{w} \rangle_{\Omega_h^0} = \sum_{M \in \Omega_h^0} u_M w_M, \qquad \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle_{\Omega_h} = \sum_{M \in \Omega_h} \alpha_M \beta_M,$$
$$\langle \boldsymbol{\varphi}, \boldsymbol{\psi} \rangle_{\Gamma_h^0} = \sum_{M \in \Gamma_h^0} \varphi_M \psi_M, \qquad \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle_{\Gamma_h} = \sum_{M \in \Gamma_h} \alpha_M \beta_M,$$
$$\left\langle \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varphi} \end{pmatrix}, \begin{pmatrix} \mathbf{w} \\ \boldsymbol{\psi} \end{pmatrix} \right\rangle_{\Omega_h^0 \times \Gamma_h^0} = \langle \mathbf{u}, \mathbf{w} \rangle_{\Omega_h^0} + \langle \boldsymbol{\varphi}, \boldsymbol{\psi} \rangle_{\Gamma_h^0}.$$

2. A diagonal mass matrix $\mathbf{B}(\boldsymbol{\alpha})$

$$\left(\mathbf{B}(\boldsymbol{\alpha})\mathbf{u}\right)_M = W_M \alpha_M u_M \quad \forall M \in \Omega_h^0$$

where W_M is a weight taking values 1 at interior nodes, 1/2 at boundary nodes and 1/4 at corner nodes.

3. A penta-diagonal stiffness matrix $\mathbf{A}(\boldsymbol{\alpha})$

$$\left(\mathbf{A}(\boldsymbol{\alpha})\mathbf{u}\right)_M = W_M \left(\frac{(\alpha_M + \alpha_R)(u_M - u_R) + (\alpha_M + \alpha_L)(u_M - u_L)}{2\Delta x^2} \right.$$
$$\left. + \frac{(\alpha_M + \alpha_U)(u_M - u_U) + (\alpha_M + \alpha_D)(u_M - u_D)}{2\Delta z^2} \right) \quad \forall M \in \Omega_h^0$$

where U (upper), R (right), D (down) and L (left) denote the neighbors of M. We use the convention that any missing value is completed by symmetry with respect to M. For example, at a point of the lower boundary, the missing u_D is replaced by an additional u_U.

And a correlation section $\mathbf{a}(.,.)$

$$\mathbf{a}(\mathbf{u},\mathbf{w})_M = W_M \left(\frac{(u_M - u_R)(w_M - w_R) + (u_M - u_L)(w_M - w_L)}{2\Delta x^2} + \frac{(u_M - u_U)(w_M - w_U) + (u_M - u_D)(w_M - w_D)}{2\Delta z^2} \right) \quad \forall M \in \Omega_h$$

with the same convention when neighbors are missing.

4. A diagonal dissipation matrix $\mathbf{D}(\boldsymbol{\alpha})$

$$(\mathbf{D}(\boldsymbol{\alpha})\mathbf{u})_M = WB_M \alpha_M u_M \quad \forall M \in \Omega_h^0$$

where WB_M is a weight taking values 0 at interior nodes, $1/\Delta x$ at vertical boundary nodes, $1/\Delta z$ at horizontal boundary nodes and $(1/\Delta x + 1/\Delta z)/2$ at corner nodes.

5. A tri-diagonal matrix $\mathbf{C}(\boldsymbol{\alpha})$

$$\begin{aligned}(\mathbf{C}(\boldsymbol{\alpha})\mathbf{u})_M =& WB'_{MP}(\alpha_M + \alpha_P)(u_M - u_P) \\ &+ WB'_{MQ}(\alpha_M + \alpha_Q)(u_M - u_Q) \quad \forall M \in \Omega_h^0\end{aligned}$$

where P and Q denote the neighbors of M on the boundary, and where WB'_{MP} is a weight taking values 0 at interior segments, $1/2\Delta x \Delta z^2$ at vertical boundary segments, and $1/2\Delta x^2 \Delta z$ at horizontal boundary segments. And a correlation section $\mathbf{c}(.,.)$

$$\begin{aligned}\mathbf{c}(\mathbf{u},\boldsymbol{\psi})_M =& WB'_{MP}(u_M - u_P)(\psi_M - \psi_P) \\ &+ WB'_{MQ}(u_M - u_Q)(\psi_M - \psi_Q) \quad \forall M \in \Gamma_h.\end{aligned}$$

6. A diagonal matrix $\mathbf{E}(\boldsymbol{\alpha})$

$$(\mathbf{E}(\boldsymbol{\alpha})\mathbf{u})_M = WC_M \alpha_M u_M \quad \forall M \in \Omega_h^0$$

where WC_M is a weight taking values 0 at interior and boundary nodes, and $3/2\Delta x \Delta z$ at corner nodes.

B.1.2 Time discretization

The observation time interval $[0,T]$ is divided into NT time steps of length Δt. For any sequence of wavefields $\mathbf{u} = (\mathbf{u}^k, k = 1,\dots,\mathsf{NT})$ and $\mathbf{w} = (\mathbf{w}^k, k = 1,\dots,\mathsf{NT})$, and for any sequence of auxiliary fields $\boldsymbol{\varphi} = (\boldsymbol{\varphi}^k, k = 1,\dots,\mathsf{NT})$ and $\boldsymbol{\psi} = (\boldsymbol{\psi}^k, k = 1,\dots,\mathsf{NT})$, we define

1. scalar products $\langle .,. \rangle_{\Omega_h^0,t}$, $\langle .,. \rangle_{\Gamma_h^0,t}$ and $\langle .,. \rangle_{\Omega_h^0 \times \Gamma_h^0,t}$

$$\langle \mathbf{u},\mathbf{w} \rangle_{\Omega_h^0,t} = \sum_{k=1}^{\mathsf{NT}} \langle \mathbf{u}^k, \mathbf{w}^k \rangle_{\Omega_h^0}, \qquad \langle \boldsymbol{\varphi},\boldsymbol{\psi} \rangle_{\Gamma_h^0,t} = \sum_{k=1}^{\mathsf{NT}} \left\langle \boldsymbol{\varphi}^k, \boldsymbol{\psi}^k \right\rangle_{\Gamma_h^0},$$

$$\left\langle \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varphi} \end{pmatrix}, \begin{pmatrix} \mathbf{w} \\ \boldsymbol{\psi} \end{pmatrix} \right\rangle_{\Omega_h^0 \times \Gamma_h^0,t} = \langle \mathbf{u},\mathbf{w} \rangle_{\Omega_h^0,t} + \langle \boldsymbol{\varphi},\boldsymbol{\psi} \rangle_{\Gamma_h^0,t},$$

2. a time correlation $\langle .,. \rangle_t$, for any $M \in \Omega_h^0$,

$$(\langle \mathbf{u},\mathbf{w} \rangle_t)_M = \sum_{k=1}^{\mathsf{NT}} u_M^k w_M^k,$$

3. the discrete second and first derivatives in time at $t^k = k\Delta t$, $k = 2,\dots,\mathsf{NT}-1$

$$\mathbf{u}^{(2)k} = \frac{\mathbf{u}^{k+1} - 2\mathbf{u}^k + \mathbf{u}^{k-1}}{\Delta t^2}, \qquad \mathbf{u}^{(1)k} = \frac{\mathbf{u}^{k+1} - \mathbf{u}^{k-1}}{2\Delta t}.$$

B.1.3 Acquisition

We denote by $\mathcal{G}$ the set of all geophone locations for the current shot. Seismograms are computed at all geophone locations $G \in \mathcal{G}$ for all time steps $k = 2, \ldots, \mathsf{NT}$, with the convention that they take the value zero at $k = 0, 1$.

For any seismograms $\mathbf{c} = (c_G^k, G \in \mathcal{G}, k = 2, \ldots, \mathsf{NT})$ and $\mathbf{d} = (d_G^k, G \in \mathcal{G}, k = 2, \ldots, \mathsf{NT})$, we define

1. a scalar product $\langle ., . \rangle_{\mathcal{G},t}$

$$\langle \mathbf{c}, \mathbf{d} \rangle_{\mathcal{G},t} = \sum_{k=2}^{\mathsf{NT}} \sum_{G \in \mathcal{G}} c_G^k d_G^k,$$

2. a semi-norm $\|.\|_{WG}$

$$\|\mathbf{c}\|_{WG}^2 = \langle \mathbf{WGc}, \mathbf{c} \rangle_{\mathcal{G},t} = \sum_{k=2}^{\mathsf{NT}} \sum_{G \in \mathcal{G}} WG_G^k (c_G^k)^2$$

where WG_G^k is a weight taking into account the mute of the direct arrivals and the correction for the geometrical spreading.

B.2 The forward wave equation

We use exactly the same scheme of second order in space and in time as in [4], but for second order absorbing boundary conditions, which are formulated in a variational way with auxiliary functions on the artificial boundaries. The fully discretized wave equation becomes

$$\begin{aligned} &\mathbf{B}(\boldsymbol{\nu}^2\boldsymbol{\mu})\mathbf{u}^{(2)k} + \mathbf{A}(\boldsymbol{\mu})\mathbf{u}^k + \mathbf{D}(\boldsymbol{\nu}\boldsymbol{\mu})\mathbf{u}^{(1)k} - \frac{1}{2}\boldsymbol{\varphi}^{(1)k} = \mathbf{f}^k \\ &\boldsymbol{\varphi}^{(2)k} + \mathbf{C}(\frac{\boldsymbol{\mu}}{\boldsymbol{\nu}})\mathbf{u}^k + \mathbf{E}(\frac{1}{\boldsymbol{\mu}})\mathbf{u}^{(1)k} = 0 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \ldots, \mathsf{NT}-1, \\ &\mathbf{u}^0 = \mathbf{u}^1 = \boldsymbol{\varphi}^0 = \boldsymbol{\varphi}^1 = 0, \end{aligned} \tag{9}$$

where $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ are the slowness and specific volume sections of the model, $\mathbf{u}^k$ and $\boldsymbol{\varphi}^k$ represent the wavefield and the auxiliary field at time t^k, and $\mathbf{f}^k$ represents the action of the source at time t^k (scaled by $1/\Delta x \Delta z$). Notice that, because of the variational approach used to establish this equation, the matrices $\mathbf{A}(\boldsymbol{\alpha})$ and $\mathbf{C}(\boldsymbol{\alpha})$ and the correlations $\mathbf{a}(.,.)$ and $\mathbf{c}(.,.)$ satisfy the identities: for any $\mathbf{u}, \mathbf{w} \in \mathbb{R}^{|\Omega_h^0|}$, for any $\boldsymbol{\psi} \in \mathbb{R}^{|\Gamma_h^0|}$ and for any $\boldsymbol{\alpha} \in \mathbb{R}^{|\Omega_h|}$,

$$\begin{aligned} \langle \mathbf{A}(\boldsymbol{\alpha})\mathbf{u}, \mathbf{w} \rangle_{\Omega_h^0} &= \langle \mathbf{u}, \mathbf{A}(\boldsymbol{\alpha})\mathbf{w} \rangle_{\Omega_h^0} = \langle \boldsymbol{\alpha}, \mathbf{a}(\mathbf{u}, \mathbf{w}) \rangle_{\Omega_h}, \\ \langle \mathbf{C}(\boldsymbol{\alpha})\mathbf{u}, \boldsymbol{\psi} \rangle_{\Gamma_h^0} &= \langle \mathbf{u}, \mathbf{C}(\boldsymbol{\alpha})\boldsymbol{\psi} \rangle_{\Gamma_h^0} = \langle \boldsymbol{\alpha}, \mathbf{c}(\mathbf{u}, \boldsymbol{\psi}) \rangle_{\Gamma_h}. \end{aligned} \tag{10}$$

B.3 The data misfit function

Let $\mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_{\mathsf{NS}})$ be the collection of data seismograms. For each shot $n = 1, \ldots, \mathsf{NS}$, we denote by $\boldsymbol{\mathcal{O}}_n$ the observation matrix which computes the pressure at all geophone locations $G \in \mathcal{G}_n$ from the wavefield $\mathbf{u}_n$. The data misfit function J_n for the nth shot becomes

$$J_n = \frac{1}{2}\|\mathbf{d}_n - \boldsymbol{\mathcal{O}}_n \mathbf{u}_n\|_{WG}^2 \tag{11}$$

and the total data misfit function J becomes

$$J = \sum_{n=1}^{\mathsf{NS}} J_n. \tag{12}$$

B.4 The gradient of the data misfit function

Once the equations for the wave equation (9) and for the total data misfit function (11,12) have been chosen, the gradient of J with respect to the parameters $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ at given smooth background slowness $\boldsymbol{\nu}_s$ and specific volume $\boldsymbol{\mu}_s$ can be computed by the following formulas. For each shot $n = 1, \ldots, \mathsf{NS}$,

1. forward propagation of the source

$$\begin{aligned}
&\mathbf{B}(\boldsymbol{\nu}_s^2\boldsymbol{\mu}_s)\mathbf{u}_{s,n}^{(2)k} + \mathbf{A}(\boldsymbol{\mu}_s)\mathbf{u}_{s,n}^{k} + \mathbf{D}(\boldsymbol{\nu}_s\boldsymbol{\mu}_s)\mathbf{u}_{s,n}^{(1)k} - \frac{1}{2}\boldsymbol{\varphi}_{s,n}^{(1)k} = \mathbf{f}_n^k \\
&\boldsymbol{\varphi}_{s,n}^{(2)k} + \mathbf{C}(\frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s})\mathbf{u}_{s,n}^{k} + \mathbf{E}(\frac{1}{\boldsymbol{\mu}_s})\mathbf{u}_{s,n}^{(1)k} = 0 \qquad k = 1, \ldots, \mathsf{NT}-1, \\
&\mathbf{u}_{s,n}^0 = \mathbf{u}_{s,n}^1 = \boldsymbol{\varphi}_{s,n}^0 = \boldsymbol{\varphi}_{s,n}^1 = 0,
\end{aligned} \tag{13}$$

2. computation of the residual

$$\mathbf{e}_{s,n}^k = \mathbf{d}_n^k - \boldsymbol{\mathcal{O}}_n\mathbf{u}_{s,n}^k \quad k = 2, \ldots, \mathsf{NT}, \tag{14}$$

3. backward propagation of the residual

$$\begin{aligned}
&\mathbf{B}(\boldsymbol{\nu}_s^2\boldsymbol{\mu}_s)\mathbf{w}_{s,n}^{(2)k} + \mathbf{A}(\boldsymbol{\mu}_s)\mathbf{w}_{s,n}^{k} - \mathbf{D}(\boldsymbol{\nu}_s\boldsymbol{\mu}_s)\mathbf{w}_{s,n}^{(1)k} \\
&\qquad + \mathbf{C}(\frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s})\boldsymbol{\psi}_{s,n}^{k} - \mathbf{E}(\frac{1}{\boldsymbol{\mu}_s})\boldsymbol{\psi}_{s,n}^{(1)k} = \boldsymbol{\mathcal{O}}_n^T\mathbf{W}\mathbf{G}_n^k\mathbf{e}_{s,n}^k \\
&\frac{1}{2}\mathbf{w}_{s,n}^{(1)k} + \boldsymbol{\psi}_{s,n}^{(2)k} = 0 \qquad k = \mathsf{NT}, \ldots, 2, \\
&\mathbf{w}_{s,n}^{\mathsf{NT}} = \mathbf{w}_{s,n}^{\mathsf{NT}-1} = \boldsymbol{\psi}_{s,n}^{\mathsf{NT}} = \boldsymbol{\psi}_{s,n}^{\mathsf{NT}-1} = 0,
\end{aligned} \tag{15}$$

4. correlations of the forward and backward wavefields

$$\begin{aligned}
\mathbf{g}_{s,n}^{\boldsymbol{\nu}} &= \left\langle 2\mathbf{B}(\boldsymbol{\nu}_s\boldsymbol{\mu}_s)\mathbf{u}_{s,n}^{(2)}, \mathbf{w}_{s,n}\right\rangle_t + \left\langle \mathbf{D}(\boldsymbol{\mu}_s)\mathbf{u}_{s,n}^{(1)}, \mathbf{w}_{s,n}\right\rangle_t - \left\langle \frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s^2}, \mathbf{c}(\mathbf{u}_{s,n}, \boldsymbol{\psi}_{s,n})\right\rangle_t, \\
\mathbf{g}_{s,n}^{\boldsymbol{\mu}} &= \left\langle \mathbf{B}(\boldsymbol{\nu}_s^2)\mathbf{u}_{s,n}^{(2)}, \mathbf{w}_{s,n}\right\rangle_t + \langle \mathbf{1}, \mathbf{a}(\mathbf{u}_{s,n}, \mathbf{w}_{s,n})\rangle_t + \left\langle \mathbf{D}(\boldsymbol{\nu}_s)\mathbf{u}_{s,n}^{(1)}, \mathbf{w}_{s,n}\right\rangle_t \\
&\quad + \left\langle \frac{1}{\boldsymbol{\nu}_s}, \mathbf{c}(\mathbf{u}_{s,n}, \boldsymbol{\psi}_{s,n})\right\rangle_t - \left\langle \mathbf{E}(\frac{1}{\boldsymbol{\mu}_s^2})\mathbf{u}_{s,n}^{(1)}, \boldsymbol{\psi}_{s,n}\right\rangle_t.
\end{aligned} \tag{16}$$

Then,

5. stack of the correlations for each shot

$$\nabla_{\boldsymbol{\nu}} J = \sum_{n=1}^{\mathsf{NS}} \mathbf{g}_{s,n}^{\boldsymbol{\nu}}, \quad \nabla_{\boldsymbol{\mu}} J = \sum_{n=1}^{\mathsf{NS}} \mathbf{g}_{s,n}^{\boldsymbol{\mu}}. \tag{17}$$

B.5 Proof of the formulas for the gradient

Firstly, we consider the case of a single shot and hence we drop the shot index n. For simplicity, we also drop the smooth index s.

1. Let $\mathsf{NST} = |\Omega_h^0|(\mathsf{NT}-1)$ and $\mathsf{NBT} = |\Gamma_h^0|(\mathsf{NT}-1)$. We define the forward wave equation operator: for any parameter fields $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon} \in \mathbb{R}^{|\Omega_h|}$

$$\begin{aligned}
\boldsymbol{\mathcal{W}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon}) : \begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varphi}\end{pmatrix} = \begin{pmatrix}\mathbf{u}^2, \ldots, \mathbf{u}^{\mathsf{NT}} \\ \boldsymbol{\varphi}^2, \ldots, \boldsymbol{\varphi}^{\mathsf{NT}}\end{pmatrix} &\in \mathbb{R}^{\mathsf{NST}} \times \mathbb{R}^{\mathsf{NBT}} \\
&\longmapsto \begin{pmatrix}\mathbf{f}_p \\ \mathbf{f}_\varphi\end{pmatrix} = \begin{pmatrix}\mathbf{f}_p^1, \ldots, \mathbf{f}_p^{\mathsf{NT}-1} \\ \mathbf{f}_\varphi^1, \ldots, \mathbf{f}_\varphi^{\mathsf{NT}-1}\end{pmatrix} \in \mathbb{R}^{\mathsf{NST}} \times \mathbb{R}^{\mathsf{NBT}},
\end{aligned}$$

where $\mathbf{f}_p$ and $\mathbf{f}_\varphi$ are defined by

$$\begin{aligned}\mathbf{f}_p^k &= \mathbf{B}(\boldsymbol{\beta})\mathbf{u}^{(2)k} + \mathbf{A}(\boldsymbol{\alpha})\mathbf{u}^k + \mathbf{D}(\boldsymbol{\delta})\mathbf{u}^{(1)k} - \frac{1}{2}\boldsymbol{\varphi}^{(1)k}\\ \mathbf{f}_\varphi^k &= \boldsymbol{\varphi}^{(2)k} + \mathbf{C}(\boldsymbol{\gamma})\mathbf{u}^k + \mathbf{E}(\varepsilon)\mathbf{u}^{(1)k}\\ &\qquad\qquad k = 1, \dots, \mathsf{NT}-1,\end{aligned}$$

with the convention that

$$\mathbf{u}^0 = \mathbf{u}^1 = \boldsymbol{\varphi}^0 = \boldsymbol{\varphi}^1 = 0.$$

With this notation, equation (13) simplifies to

$$\boldsymbol{\mathcal{W}}\left(\boldsymbol{\nu}^2\boldsymbol{\mu}, \boldsymbol{\mu}, \boldsymbol{\nu}\boldsymbol{\mu}, \frac{\boldsymbol{\mu}}{\boldsymbol{\nu}}, \frac{1}{\boldsymbol{\mu}}\right)\begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varphi}\end{pmatrix} = \begin{pmatrix}\mathbf{f}\\ \mathbf{0}\end{pmatrix}.$$

2. The transposed matrix $\boldsymbol{\mathcal{W}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \varepsilon)^T$ is defined by: for any $\mathbf{u} = (\mathbf{u}^2, \dots, \mathbf{u}^{\mathsf{NT}}) \in \mathbb{R}^{\mathsf{NST}}$, $\boldsymbol{\varphi} = (\boldsymbol{\varphi}^2, \dots, \boldsymbol{\varphi}^{\mathsf{NT}}) \in \mathbb{R}^{\mathsf{NBT}}$, $\mathbf{w} = (\mathbf{w}^1, \dots, \mathbf{w}^{\mathsf{NT}-1}) \in \mathbb{R}^{\mathsf{NST}}$ and $\boldsymbol{\psi} = (\boldsymbol{\psi}^1, \dots, \boldsymbol{\psi}^{\mathsf{NT}-1}) \in \mathbb{R}^{\mathsf{NBT}}$,

$$\begin{aligned}&\left\langle \boldsymbol{\mathcal{W}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \varepsilon)\begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varphi}\end{pmatrix}, \begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}\end{pmatrix}\right\rangle_{\Omega_h^0 \times \Gamma_h^0, t}\\ &\qquad = \left\langle \begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varphi}\end{pmatrix}, \boldsymbol{\mathcal{W}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \varepsilon)^T\begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}\end{pmatrix}\right\rangle_{\Omega_h^0 \times \Gamma_h^0, t}. \quad (18)\end{aligned}$$

It is easy to see by performing discrete part-by-part integrations and using the fact that matrices $\mathbf{B}(\boldsymbol{\beta})$, $\mathbf{A}(\boldsymbol{\alpha})$, $\mathbf{D}(\boldsymbol{\delta})$, $\mathbf{C}(\boldsymbol{\gamma})$ and $\mathbf{E}(\varepsilon)$ are symmetrical that $\boldsymbol{\mathcal{W}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \varepsilon)^T$ is the backward wave equation operator: for any parameter fields $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon} \in \mathbb{R}^{|\Omega_h|}$,

$$\begin{aligned}\boldsymbol{\mathcal{W}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \varepsilon)^T : \begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}\end{pmatrix} &= \begin{pmatrix}\mathbf{w}^1, \dots, \mathbf{w}^{\mathsf{NT}-1}\\ \boldsymbol{\psi}^1, \dots, \boldsymbol{\psi}^{\mathsf{NT}-1}\end{pmatrix} \in \mathbb{R}^{\mathsf{NST}} \times \mathbb{R}^{\mathsf{NBT}}\\ &\longmapsto \begin{pmatrix}\mathbf{g}_q\\ \mathbf{g}_\psi\end{pmatrix} = \begin{pmatrix}\mathbf{g}_q^2, \dots, \mathbf{g}_q^{\mathsf{NT}}\\ \mathbf{g}_\psi^2, \dots, \mathbf{g}_\psi^{\mathsf{NT}}\end{pmatrix} \in \mathbb{R}^{\mathsf{NST}} \times \mathbb{R}^{\mathsf{NBT}},\end{aligned}$$

where $\mathbf{g}_q$ and $\mathbf{g}_\psi$ are defined by

$$\begin{aligned}\mathbf{g}_q^k &= \mathbf{B}(\boldsymbol{\beta})\mathbf{w}^{(2)k} + \mathbf{A}(\boldsymbol{\alpha})\mathbf{w}^k - \mathbf{D}(\boldsymbol{\delta})\mathbf{w}^{(1)k} + \mathbf{C}(\boldsymbol{\gamma})\boldsymbol{\psi}^k - \mathbf{E}(\varepsilon)\boldsymbol{\psi}^{(1)k},\\ \mathbf{g}_\psi^k &= \tfrac{1}{2}\mathbf{w}^{(1)k} + \boldsymbol{\psi}^{(2)k}\\ &\qquad\qquad k = \mathsf{NT}, \dots, 2,\end{aligned} \quad (19)$$

with the convention that

$$\mathbf{w}^{\mathsf{NT}+1} = \mathbf{w}^{\mathsf{NT}} = \boldsymbol{\psi}^{\mathsf{NT}+1} = \boldsymbol{\psi}^{\mathsf{NT}} = 0.$$

Notice that when the second part of the right-hand side $\mathbf{g}_\psi$ is equal to zero, (19) is equivalent to

$$\begin{aligned}\mathbf{g}_q^k &= \mathbf{B}(\boldsymbol{\beta})\mathbf{w}^{(2)k} + \mathbf{A}(\boldsymbol{\alpha})\mathbf{w}^k - \mathbf{D}(\boldsymbol{\delta})\mathbf{w}^{(1)k} + \tfrac{1}{2}\boldsymbol{\Psi}^{(1)k}\\ 0 = \mathbf{g}_\psi^k &= \boldsymbol{\Psi}^{(2)k} + \mathbf{C}(\boldsymbol{\gamma})\mathbf{w} - \mathbf{E}(\varepsilon)\mathbf{w}^{(1)k}\\ &\qquad\qquad k = \mathsf{NT}, \dots, 2,\end{aligned}$$

with the same convention that

$$\boldsymbol{\Psi}^{\mathsf{NT}+1} = \boldsymbol{\Psi}^{\mathsf{NT}} = 0.$$

3. The following identity, which results from equations (10), (18) and discrete part-by-part integration, will be most useful in the sequel: for any $\mathbf{u}, \mathbf{w} \in \mathbb{R}^{\mathsf{NST}}$, for any $\boldsymbol{\varphi}, \boldsymbol{\psi} \in \mathbb{R}^{\mathsf{NBT}}$ and for any $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon} \in \mathbb{R}^{|\Omega_h|}$,

$$\begin{aligned}\left\langle \boldsymbol{\mathcal{W}}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\varepsilon}) \begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varphi}\end{pmatrix}, \begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}\end{pmatrix} \right\rangle_{\Omega_h^0\times\Gamma_h^0,t} &= \left\langle \begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varphi}\end{pmatrix}, \boldsymbol{\mathcal{W}}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\varepsilon})^T \begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}\end{pmatrix} \right\rangle_{\Omega_h^0\times\Gamma_h^0,t} \\ &= \aleph_1+\aleph_2+\aleph_3+\aleph_4+\aleph_5+\aleph_6+\aleph_7, \quad (20)\end{aligned}$$

where

$$\begin{aligned}
\aleph_1 &= \left\langle \mathbf{B}(\boldsymbol{\beta})\mathbf{u}^{(2)}, \mathbf{w} \right\rangle_{\Omega_h^0,t} = \left\langle \mathbf{u}, \mathbf{B}(\boldsymbol{\beta})\mathbf{w}^{(2)} \right\rangle_{\Omega_h^0,t} \\
\aleph_2 &= \langle \mathbf{A}(\boldsymbol{\alpha})\mathbf{u}, \mathbf{w}\rangle_{\Omega_h^0,t} = \langle \boldsymbol{\alpha}, \mathbf{a}(\mathbf{u},\mathbf{w})\rangle_{\Omega_h,t} = \langle \mathbf{u}, \mathbf{A}(\boldsymbol{\alpha})\mathbf{w}\rangle_{\Omega_h^0,t} \\
\aleph_3 &= \left\langle \mathbf{D}(\boldsymbol{\delta})\mathbf{u}^{(1)}, \mathbf{w} \right\rangle_{\Omega_h^0,t} = -\left\langle \mathbf{u}, \mathbf{D}(\boldsymbol{\delta})\mathbf{w}^{(1)} \right\rangle_{\Omega_h^0,t} \\
\aleph_4 &= -\frac{1}{2}\left\langle \boldsymbol{\varphi}^{(1)}, \mathbf{w} \right\rangle_{\Omega_h^0,t} = \frac{1}{2}\left\langle \boldsymbol{\varphi}, \mathbf{w}^{(1)} \right\rangle_{\Omega_h^0,t} \\
\aleph_5 &= \left\langle \boldsymbol{\varphi}^{(2)}, \boldsymbol{\psi} \right\rangle_{\Gamma_h^0,t} = \left\langle \boldsymbol{\varphi}, \boldsymbol{\psi}^{(2)} \right\rangle_{\Gamma_h^0,t} \\
\aleph_6 &= \langle \mathbf{C}(\boldsymbol{\gamma})\mathbf{u}, \boldsymbol{\psi}\rangle_{\Gamma_h^0,t} = \langle \boldsymbol{\gamma}, \mathbf{c}(\mathbf{u},\boldsymbol{\psi})\rangle_{\Gamma_h,t} = \langle \mathbf{u}, \mathbf{C}(\boldsymbol{\gamma})\boldsymbol{\psi}\rangle_{\Gamma_h^0,t} \\
\aleph_7 &= \left\langle \mathbf{E}(\boldsymbol{\varepsilon})\mathbf{u}^{(1)}, \boldsymbol{\psi} \right\rangle_{\Gamma_h^0,t} = -\left\langle \mathbf{u}, \mathbf{E}(\boldsymbol{\varepsilon})\boldsymbol{\psi}^{(1)} \right\rangle_{\Gamma_h^0,t}
\end{aligned}$$

4. The Lagrangian function associated to the problem of minimizing the data misfit function defined by (11), under the constraint that the forward wave equation (13) is satisfied, can be written: for any wavefields $\mathbf{u}, \mathbf{w} \in \mathbb{R}^{\mathsf{NST}}$, for any auxiliary fields $\boldsymbol{\varphi}, \boldsymbol{\psi} \in \mathbb{R}^{\mathsf{NBT}}$, and for any parameter fields $\boldsymbol{\nu}, \boldsymbol{\mu} \in \mathbb{R}^{|\Omega_h|}$,

$$\begin{aligned}L(\mathbf{u},\boldsymbol{\varphi},\mathbf{w},\boldsymbol{\psi};\boldsymbol{\nu},\boldsymbol{\mu}) = \frac{1}{2}\langle \mathbf{WG}(\mathbf{d}-\boldsymbol{\mathcal{O}}\mathbf{u}), \mathbf{d}-\boldsymbol{\mathcal{O}}\mathbf{u}\rangle_{G,t} \\ + \left\langle \boldsymbol{\mathcal{W}}\left(\boldsymbol{\nu}^2\boldsymbol{\mu}, \boldsymbol{\mu}, \boldsymbol{\nu}\boldsymbol{\mu}, \frac{\boldsymbol{\mu}}{\boldsymbol{\nu}}, \frac{1}{\boldsymbol{\mu}}\right)\begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varphi}\end{pmatrix} - \begin{pmatrix}\mathbf{f}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}\end{pmatrix} \right\rangle_{\Omega_h^0\times\Gamma_h^0,t}.\end{aligned}$$

For any $\boldsymbol{\nu}, \boldsymbol{\mu}$, if $\mathbf{u}$ and $\boldsymbol{\varphi}$ are solution of the forward wave equation (13), then the Lagrangian reduces to the data misfit function, and moreover, if $\mathbf{w}$ and $\boldsymbol{\psi}$ are chosen so as to satisfy the adjoint equation

$$\boldsymbol{\mathcal{W}}\left(\boldsymbol{\nu}^2\boldsymbol{\mu}, \boldsymbol{\mu}, \boldsymbol{\nu}\boldsymbol{\mu}, \frac{\boldsymbol{\mu}}{\boldsymbol{\nu}}, \frac{1}{\boldsymbol{\mu}}\right)^T \begin{pmatrix}\mathbf{w}\\ \boldsymbol{\psi}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mathcal{O}}^T\mathbf{WG}(\mathbf{d}-\boldsymbol{\mathcal{O}}\mathbf{u})\\ \mathbf{0}\end{pmatrix}, \quad (21)$$

then the formal differentiation of the Lagrangian reduces to

$$\delta L = \delta J = \frac{\partial L}{\partial \boldsymbol{\nu}}(\mathbf{u},\boldsymbol{\varphi},\mathbf{w},\boldsymbol{\psi};\boldsymbol{\nu},\boldsymbol{\mu})\delta\boldsymbol{\nu} + \frac{\partial L}{\partial \boldsymbol{\mu}}(\mathbf{u},\boldsymbol{\varphi},\mathbf{w},\boldsymbol{\psi};\boldsymbol{\nu},\boldsymbol{\mu})\delta\boldsymbol{\mu}.$$

Using equation (19), we see that (21) is nothing but the equations (14) and (15) for the backward propagation of the residual.

We define a correlation operator: for any parameter fields $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon} \in \mathbf{R}^{|\Omega_h|}$, for any $\mathbf{u} = (\mathbf{u}^2, \ldots, \mathbf{u}^{\mathsf{NT}}) \in \mathbf{R}^{\mathsf{NST}}$, $\mathbf{w} = (\mathbf{w}^1, \ldots, \mathbf{w}^{\mathsf{NT}-1}) \in \mathbf{R}^{\mathsf{NST}}$ and $\boldsymbol{\psi} = (\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^{\mathsf{NT}-1}) \in \mathbf{R}^{\mathsf{NBT}}$

$$\boldsymbol{\chi}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon}) \left(\begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{w} \\ \boldsymbol{\psi} \end{pmatrix} \right) = \left\langle \mathbf{B}(\boldsymbol{\beta})\mathbf{u}^{(2)}, \mathbf{w} \right\rangle_t + \langle \boldsymbol{\alpha}, \mathbf{a}(\mathbf{u}, \mathbf{w}) \rangle_t + \left\langle \mathbf{D}(\boldsymbol{\delta})\mathbf{u}^{(1)}, \mathbf{w} \right\rangle_t + \langle \boldsymbol{\gamma}, \mathbf{c}(\mathbf{u}, \boldsymbol{\psi}) \rangle_t + \left\langle \mathbf{E}(\boldsymbol{\varepsilon})\mathbf{u}^{(1)}, \boldsymbol{\psi} \right\rangle_t .$$

The gradient sections $\nabla_{\boldsymbol{\nu}} J$ and $\nabla_{\boldsymbol{\mu}} J$ are determined by: for any parameter fields $\delta\boldsymbol{\nu}, \delta\boldsymbol{\mu} \in \mathbf{R}^{|\Omega_h|}$,

$$\langle \nabla_{\boldsymbol{\nu}} J, \delta\boldsymbol{\nu} \rangle_{\Omega_h} = \left\langle \boldsymbol{\mathcal{W}} \left(2\boldsymbol{\nu}\boldsymbol{\mu}\delta\boldsymbol{\nu}, \mathbf{0}, \boldsymbol{\mu}\delta\boldsymbol{\nu}, -\frac{\boldsymbol{\mu}\delta\boldsymbol{\nu}}{\boldsymbol{\nu}^2}, \mathbf{0} \right) \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{w} \\ \boldsymbol{\psi} \end{pmatrix} \right\rangle_{\Omega_h^0 \times \Gamma_h^0, t},$$

$$\langle \nabla_{\boldsymbol{\mu}} J, \delta\boldsymbol{\mu} \rangle_{\Omega_h} = \left\langle \boldsymbol{\mathcal{W}} \left(\boldsymbol{\nu}^2\delta\boldsymbol{\mu}, \delta\boldsymbol{\mu}, \boldsymbol{\nu}\delta\boldsymbol{\mu}, -\frac{\delta\boldsymbol{\mu}}{\boldsymbol{\nu}}, -\frac{\delta\boldsymbol{\mu}}{\boldsymbol{\mu}^2} \right) \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{w} \\ \boldsymbol{\psi} \end{pmatrix} \right\rangle_{\Omega_h^0 \times \Gamma_h^0, t},$$

which immediately gives the formulas (16) by using the identity (20) and picking up the coefficients of $\delta\boldsymbol{\nu}$ and $\delta\boldsymbol{\mu}$.

Then, we consider all the shots and formulas (17) immediately come from (12).

C Multiscale representation

We first construct a family of smooth/rough decompositions, then we describe the computation of the components on the local basis from the components on the multiscale basis.

C.1 Multiscale representation

We suppose here that $\mathsf{NX} = \mathsf{NX}_0 2^{\mathsf{NSCALE}}$ and $\mathsf{NZ} = \mathsf{NZ}_0 2^{\mathsf{NSCALE}}$, where NX_0 and NZ_0 are small numbers. By dichotomy on the number of cells in each direction, we define an embedded sequence of regular grids

$$\Omega_0 \subset \ldots \subset \Omega_{\mathsf{NSCALE}} = \Omega_h.$$

We set $\Sigma_0 = \Omega_0$ and for all scale $k = 1, \ldots, \mathsf{NSCALE}$, we denote by Σ_k the complement of Ω_{k-1} in Ω_k. Hence, we have the following partition

$$\Omega_k = \Sigma_0 \cup \Sigma_1 \cup \ldots \cup \Sigma_k.$$

For all scale k, we define the space V_k of dimension $(\mathsf{NX}_0 2^k + 1)(\mathsf{NZ}_0 2^k + 1)$ of functions obtained by bilinear interpolation at the nodes of Ω_k. Each space is provided with its canonical basis of "hat" functions (which take value 1 at a given node and zero at all other nodes), denoted the local basis at scale k. Thus, we define an embedded sequence of spaces

$$V_0 \subset \ldots \subset V_{\mathsf{NSCALE}} = V.$$

We set $W_0 = V_0$ and for all scale $k = 1, \ldots, \mathsf{NSCALE}$, we denote by W_k the canonical (oblique) supplement of V_{k-1} in V_k. Hence, we have the following multiscale representation

$$V_k = W_0 \oplus W_1 \oplus \ldots \oplus W_k.$$

C.2 Smooth/rough decomposition

The smooth space is chosen as the space of the highest scale $k = \mathsf{CURSOR}$ for which property 1 is satisfied. It corresponds generally to a coarse grid with a mesh of about the size of the dominant wavelength of the signal: for the Synclay model (Appendix A), $H = 42$ m and $\bar{\lambda} = 68$ m.

The rough space is chosen as the orthogonal supplementary of V_s in V for the L^2-norm.

Notice that the smooth space can be provided with a multiscale basis which can be useful for optimization, as hierarchical optimization guaranty convergence towards the "smoothest" solution (this acts as regularization). Notice also that orthogonalization at each scale of this multiscale basis would have been very costfull.

C.3 Computation

We denote by $(\phi_M^k)_{M\in\Omega_k}$ the local basis at scale k, by $\Sigma = \Sigma_{\mathsf{CURSOR}+1} \cup \ldots \cup \Sigma_{\mathsf{NSCALE}}$ the complement of Ω_{CURSOR} in Ω_{NSCALE} and by $(\omega_M)_{M\in\Sigma}$ a basis of $V_r = V_{\mathsf{CURSOR}}^{\perp}$.

We set

$$\mathbf{B} = \left(\langle \phi_M^{\mathsf{CURSOR}}, \phi_P^{\mathsf{NSCALE}} \rangle_{\Omega_{\mathsf{NSCALE}}}\right)_{M\in\Omega_{\mathsf{CURSOR}}, P\in\Sigma}.$$

It is a rectangular matrix of $|\Omega_{\mathsf{CURSOR}}|$ lines and $|\Sigma|$ columns. We set

$$\mathbf{A} = \left(\langle \phi_M^{\mathsf{CURSOR}}, \phi_P^{\mathsf{CURSOR}} \rangle_{\Omega_{\mathsf{NSCALE}}}\right)_{M,P\in\Omega_{\mathsf{CURSOR}}}.$$

It is a positive symmetric matrix, tensor product of two tridiagonal matrices. We set

$$\mathbf{L} = \left(\langle \phi_M^{\mathsf{NSCALE}}, \phi_P^k \rangle_{\Omega_{\mathsf{NSCALE}}}\right)_{M\in\Omega_{\mathsf{CURSOR}}, (P\in\Sigma_k)_{k=0,\ldots,\mathsf{CURSOR}}}.$$

It is a lower triangular block-matrix with identity blocks on the diagonal.

Let $\mathcal{C}$ be the matrix to compute the components in the local basis at scale NSCALE from the components in the multiscale basis,

$$\mathcal{C} = \begin{pmatrix} \mathbf{1}_{\Omega_{\mathsf{CURSOR}}} & \mathbf{0} \\ \mathbf{B}^T & \mathbf{1}_\Sigma \end{pmatrix} \begin{pmatrix} \mathbf{1}_{\Omega_{\mathsf{CURSOR}}} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{1}_\Sigma \end{pmatrix} \begin{pmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_\Sigma \end{pmatrix} \tag{22}$$

and

$$\mathcal{C}^{-1} = \begin{pmatrix} \mathbf{L}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_\Sigma \end{pmatrix} \begin{pmatrix} \mathbf{1}_{\Omega_{\mathsf{CURSOR}}} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{1}_\Sigma \end{pmatrix} \begin{pmatrix} \mathbf{1}_{\Omega_{\mathsf{CURSOR}}} & \mathbf{0} \\ -\mathbf{B}^T & \mathbf{1}_\Sigma \end{pmatrix} \tag{23}$$

$$\mathcal{C}^T = \begin{pmatrix} \mathbf{L}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_\Sigma \end{pmatrix} \begin{pmatrix} \mathbf{1}_{\Omega_{\mathsf{CURSOR}}} & \mathbf{0} \\ -\mathbf{B}^T\mathbf{A}^{-1} & \mathbf{1}_\Sigma \end{pmatrix} \begin{pmatrix} \mathbf{1}_{\Omega_{\mathsf{CURSOR}}} & \mathbf{B} \\ \mathbf{0} & \mathbf{1}_\Sigma \end{pmatrix} \tag{24}$$

C.4 Applications

Let $\boldsymbol{\nu}$, $\mathbf{p}$ and $\mathbf{r}$ respectively denote the components of the slowness ν in the local basis of V, the components of the smooth slowness in the multiscale basis of V_s, and the components of the rough slowness in the basis of V_r. Then, we have the following identities:

$$\boldsymbol{\nu} = \mathcal{C}(\mathbf{P}_s^T\mathbf{p} + \mathbf{P}_r^T\mathbf{r}) \qquad \begin{cases} \mathbf{p} = \mathbf{P}_s\mathcal{C}^{-1}\boldsymbol{\nu} \\ \mathbf{r} = \mathbf{P}_r\mathcal{C}^{-1}\boldsymbol{\nu} \end{cases} \tag{25}$$

$$\nabla_{\mathbf{p}} J = \mathbf{P}_s\mathcal{C}^T\nabla_{\boldsymbol{\nu}} J \qquad \nabla_{\mathbf{r}} J = \mathbf{P}_r\mathcal{C}^T\nabla_{\boldsymbol{\nu}} J. \tag{26}$$

D The resimulation and its adjoint

We describe, as concise as possible, the fully discretized equations we use to compute the resimulation of a collection of seismograms and the gradient of the corresponding data misfit function. For simplicity, we have supposed the weight WM independent of the propagator p.

D.1 Resimulation

Given a propagator p and a time reflectivity s, the resimulation can be computed by the following formulas.

0. computation of the smooth background slowness

$$\boldsymbol{\nu}_s = \mathcal{C}\mathbf{P}_s^T\mathbf{p}. \tag{27}$$

For each shot $n = 1, \ldots, \mathsf{NS}$,

1. forward propagation of the source

$$\mathcal{W}\left(\boldsymbol{\nu}_s^2\boldsymbol{\mu}_s, \boldsymbol{\mu}_s, \boldsymbol{\nu}_s\boldsymbol{\mu}_s, \frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s}, \frac{1}{\boldsymbol{\mu}_s}\right)\begin{pmatrix}\mathbf{u}_{s,n}\\ \boldsymbol{\varphi}_{s,n}\end{pmatrix} = \begin{pmatrix}\mathbf{f}_n\\ \mathbf{0}\end{pmatrix}, \tag{28}$$

2. computation of the residual

$$\mathbf{e}_{s,n} = \mathbf{d}_n - \mathcal{O}_n\mathbf{u}_{s,n}, \tag{29}$$

3. backward propagation of the time reflectivity

$$\mathcal{W}\left(\boldsymbol{\nu}_s^2\boldsymbol{\mu}_s, \boldsymbol{\mu}_s, \boldsymbol{\nu}_s\boldsymbol{\mu}_s, \frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s}, \frac{1}{\boldsymbol{\mu}_s}\right)^T\begin{pmatrix}\mathbf{w}_{s,n}\\ \boldsymbol{\psi}_{s,n}\end{pmatrix} = \begin{pmatrix}\mathcal{O}_n^T\mathbf{W}\mathbf{G}_n\mathbf{s}_n\\ \mathbf{0}\end{pmatrix}, \tag{30}$$

4. correlation of the forward and backward wavefields

$$\mathbf{m}_{s,n} = \boldsymbol{\chi}\left(2\boldsymbol{\nu}_s\boldsymbol{\mu}_s, \mathbf{0}, \boldsymbol{\mu}_s, -\frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s^2}, \mathbf{0}\right)\left(\begin{pmatrix}\mathbf{u}_{s,n}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{w}_{s,n}\\ \boldsymbol{\psi}_{s,n}\end{pmatrix}\right). \tag{31}$$

Then,

5. stack of the correlations for each shot

$$\mathbf{m}_s = \sum_{n=1}^{\mathsf{NS}}\mathbf{m}_{s,n}, \tag{32}$$

6. quantitative estimation of the rough slowness

$$\mathbf{r} = -\mathbf{W}\mathbf{M}\mathbf{P}_r\mathcal{C}^T\mathbf{m}_s, \tag{33}$$

7. update of the slowness

$$\boldsymbol{\nu} = \boldsymbol{\nu}_s + \mathcal{C}\mathbf{P}_r^T\mathbf{r}. \tag{34}$$

For each shot $n = 1, \ldots, \mathsf{NS}$,

8. forward propagation of the source

$$\mathcal{W}\left(\boldsymbol{\nu}^2\boldsymbol{\mu}_s, \boldsymbol{\mu}_s, \boldsymbol{\nu}\boldsymbol{\mu}_s, \frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}}, \frac{1}{\boldsymbol{\mu}_s}\right)\begin{pmatrix}\mathbf{u}_n\\ \boldsymbol{\varphi}_n\end{pmatrix} = \begin{pmatrix}\mathbf{f}_n\\ \mathbf{0}\end{pmatrix}, \tag{35}$$

9. computation of the residual

$$\mathbf{e}_n = \mathbf{d}_n - \mathcal{O}_n\mathbf{u}_n. \tag{36}$$

D.2 Gradient

Given the previous computation of the resimulation, the gradient of the data misfit function with respect to the propagator p and to the time reflectivity s can be computed by the following formulas. For each shot $n = 1, \ldots, \mathsf{NS}$,

10. backward propagation of the residual

$$\boldsymbol{\mathcal{W}}\left(\boldsymbol{\nu}^2\boldsymbol{\mu}_s, \boldsymbol{\mu}_s, \boldsymbol{\nu}\boldsymbol{\mu}_s, \frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}}, \frac{1}{\boldsymbol{\mu}_s}\right)^T \begin{pmatrix}\mathbf{w}_n \\ \boldsymbol{\psi}_n\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mathcal{O}}_n^T \mathbf{W}\mathbf{G}_n\mathbf{e}_n \\ \mathbf{0}\end{pmatrix}, \tag{37}$$

11. correlation of the forward and backward wavefields

$$\mathbf{m}_n = \boldsymbol{\chi}\left(2\boldsymbol{\nu}\boldsymbol{\mu}_s, \mathbf{0}, \boldsymbol{\mu}_s, -\frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}^2}, \mathbf{0}\right)\left(\begin{pmatrix}\mathbf{u}_n \\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{w}_n \\ \boldsymbol{\psi}_n\end{pmatrix}\right). \tag{38}$$

Then,

12. stack of the correlations for each shot

$$\mathbf{m} = \sum_{n=1}^{\mathsf{NS}} \mathbf{m}_n, \tag{39}$$

13. adjoint update

$$\bar{\mathbf{r}} = \mathbf{P}_r \boldsymbol{\mathcal{C}}^T \mathbf{m}, \tag{40}$$

14. adjoint quantitative estimation of the rough slowness

$$\bar{\mathbf{m}}_s = -\boldsymbol{\mathcal{C}}\mathbf{P}_r^T \mathbf{W}\mathbf{M}\bar{\mathbf{r}}, \tag{41}$$

15. adjoint correlation

$$\bar{\mathbf{m}}_s^{\chi} = \sum_{n=1}^{\mathsf{NS}} \boldsymbol{\chi}\left(2\boldsymbol{\mu}_s\bar{\mathbf{m}}_s, \mathbf{0}, \mathbf{0}, \frac{2\boldsymbol{\mu}_s\bar{\mathbf{m}}_s}{\boldsymbol{\nu}^3}, \mathbf{0}\right)\left(\begin{pmatrix}\mathbf{u}_n \\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{w}_n \\ \boldsymbol{\psi}_n\end{pmatrix}\right). \tag{42}$$

For each shot $n = 1, \ldots, \mathsf{NS}$,

16. adjoint backward propagation

$$\boldsymbol{\mathcal{W}}\left(\boldsymbol{\nu}_s^2\boldsymbol{\mu}_s, \boldsymbol{\mu}_s, \boldsymbol{\nu}_s\boldsymbol{\mu}_s, \frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s}, \frac{1}{\boldsymbol{\mu}_s}\right)\begin{pmatrix}\bar{\mathbf{w}}_{s,n} \\ \bar{\boldsymbol{\psi}}_{s,n}\end{pmatrix} = \begin{pmatrix}-\mathbf{B}(2\boldsymbol{\nu}_s\boldsymbol{\mu}_s\bar{\mathbf{m}}_s)\mathbf{u}_{s,n}^{(2)} - \mathbf{B}(\boldsymbol{\mu}_s\bar{\mathbf{m}}_s)\mathbf{u}_{s,n}^{(1)} \\ \mathbf{C}\left(\frac{\boldsymbol{\mu}_s\bar{\mathbf{m}}_s}{\boldsymbol{\nu}_s^2}\right)\mathbf{u}_{s,n}\end{pmatrix}, \tag{43}$$

17. gradient with respect to the time reflectivity

$$\nabla_{s_n} J = -\mathbf{W}\mathbf{G}_n\boldsymbol{\mathcal{O}}_n\bar{\mathbf{w}}_{s,n}, \tag{44}$$

18. correlation of the forward and backward wavefields

$$\bar{\mathbf{m}}_{s,n}^{w} = \boldsymbol{\chi}\left(2\boldsymbol{\nu}_s\boldsymbol{\mu}_s, \mathbf{0}, \boldsymbol{\mu}_s, -\frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s^2}, \mathbf{0}\right)\left(\begin{pmatrix}\bar{\mathbf{w}}_{s,n} \\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{w}_{s,n} \\ \boldsymbol{\psi}_{s,n}\end{pmatrix}\right), \tag{45}$$

19. adjoint forward propagation

$$\boldsymbol{\mathcal{W}}\left(\boldsymbol{\nu}_s^2\boldsymbol{\mu}_s, \boldsymbol{\mu}_s, \boldsymbol{\nu}_s\boldsymbol{\mu}_s, \frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s}, \frac{1}{\boldsymbol{\mu}_s}\right)^T \begin{pmatrix}\bar{\mathbf{u}}_{s,n} \\ \bar{\boldsymbol{\varphi}}_{s,n}\end{pmatrix}$$
$$= \begin{pmatrix}-\mathbf{B}(2\boldsymbol{\nu}_s\boldsymbol{\mu}_s\bar{\mathbf{m}}_s)\mathbf{w}_{s,n}^{(2)} + \mathbf{B}(\boldsymbol{\mu}_s\bar{\mathbf{m}}_s)\mathbf{w}_{s,n}^{(1)} + \mathbf{C}\left(\frac{\boldsymbol{\mu}_s\bar{\mathbf{m}}_s}{\boldsymbol{\nu}_s^2}\right)\boldsymbol{\psi}_{s,n} \\ \mathbf{0}\end{pmatrix}, \tag{46}$$

20. correlation of the forward and backward wavefields

$$\bar{\mathbf{m}}^u_{s,n} = \boldsymbol{\chi}\left(2\boldsymbol{\nu}_s\boldsymbol{\mu}_s, \mathbf{0}, \boldsymbol{\mu}_s, -\frac{\boldsymbol{\mu}_s}{\boldsymbol{\nu}_s^2}, \mathbf{0}\right)\left(\begin{pmatrix}\mathbf{u}_{s,n}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\bar{\mathbf{u}}_{s,n}\\ \bar{\boldsymbol{\varphi}}_{s,n}\end{pmatrix}\right). \tag{47}$$

Then,

20. stack of the correlations for each shot

$$\bar{\mathbf{m}}^w_s = \sum_{n=1}^{\mathsf{NS}} \bar{\mathbf{m}}^w_{s,n}, \qquad \bar{\mathbf{m}}^u_s = \sum_{n=1}^{\mathsf{NS}} \bar{\mathbf{m}}^u_{s,n}, \tag{48}$$

21. gradient with respect to the propagator

$$\nabla_p J = \mathbf{P}_s \boldsymbol{\mathcal{C}}^T(\mathbf{m} + \bar{\mathbf{m}}^\chi_s + \bar{\mathbf{m}}^w_s + \bar{\mathbf{m}}^u_s). \tag{49}$$

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399