Parsimonious slope tomography based on eikonal solvers and the adjoint-state method

S Sambolian, S. Operto, A. Ribodetti, B. Tavakoli f., J. Virieux

HAL Id: hal-02134584
https://hal.science/hal-02134584

Parsimonious slope tomography based on eikonal solvers and the adjoint-state method

S. Sambolian [1] S. Operto,[1] A. Ribodetti,[1] B. Tavakoli F.[1,*] and J. Virieux[2]

[1]*Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, IRD, Géoazur, Valbonne, France. E-mail: sambolian@geoazur.unice.fr*
[2]*Université Grenoble Alpes, ISTerre, Grenoble, France*

SUMMARY
Velocity macromodel building is an essential step of the seismic imaging workflow. Indeed, obtaining acceptable results through migration or full waveform inversion is highly dependent on the kinematic accuracy of the background/initial velocity model. Two decades ago, stereotomography was proposed as an alternative to reflection traveltime tomography, the first relying on semi-automatic picking of locally coherent events associated with small reflection or diffraction segments tied to scatterers in depth by a pair of rays, while the latter on interpretive picking of laterally continuous reflections. The flexibility of stereotomography paved the way for many developments that have shown the efficiency of the method whilst emphasizing on the complementary information carried out by traveltimes and slopes of locally coherent events. A recent formulation recast stereotomography under a matrix-free formulation based on eikonal solvers and the adjoint-state method. In the latter, like in the previous works, the scatterer positions and the velocity field are updated jointly to tackle the ill-famed velocity–position coupling in reflection tomography. Following on from this adjoint-state formulation, we propose a new parsimonious formulation of slope tomography that offers the chance to restrain the problem to minimizing the residuals of a single data class being a slope, in search of a sole parameter class being the subsurface velocity field. This parsimonious formulation results from a variable projection, which is implemented by enforcing a consistency between the scatterer coordinates and the velocity macromodel through migration of kinematic attributes. We explain why the resulting reduced-parametrization inversion is more suitable for tomographic problems than the most common joint inversion strategy. We benchmark our method against the complex Marmousi model along with a validation through time domain full waveform inversion and then present the results of a field data case study.

Key words: Inverse theory; Tomography; Seismic velocities.

1 INTRODUCTION

The key purpose of seismic imaging methods is the retrieval of the subsurface properties like for instance wave speeds, density, attenuation or anisotropy. One of the most crucial yet challenging task in seismic imaging is velocity macromodel building. Indeed, building a kinematically accurate smooth velocity model of the subsurface from acquired seismic data is essential for obtaining reliable depth-migrated images (Etgen *et al.* 2009) or adequate starting models for full-waveform inversion (FWI; Tarantola 1984; Virieux & Operto 2009).

Several authors addressed this ill-posed inverse problem; most adopted the asymptotic high-frequency approximation (Červený 2001) whilst utilizing different types of data and methods. Initially, kinematic attributes like first-arrival traveltimes were extracted from the data and exploited (Aki & Lee 1976). The non-uniqueness of the solution in first-arrival traveltime tomography motivated the use of additional kinematic attributes as in reflection tomography (Bishop *et al.* 1985; Farra & Madariaga 1988) or even higher order attributes relying on directional reception as in polarization tomography (Hu *et al.* 1994; Farra & Le Bégat 1995) where the wholeness of the slowness vector is exploited in a transmission regime. In reflection settings, Controlled Directional Reception (CDR; Rieber 1936; Riabinkin 1957; Sword 1987), relying on locally coherent events defined by traveltime and its first-order derivative, proposed an interesting approach in the sense that
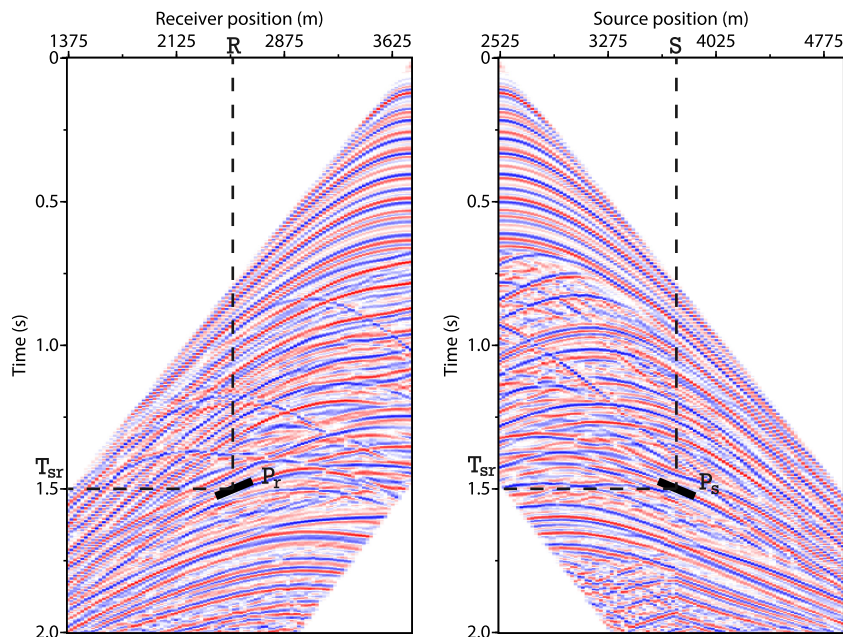
* Now at Opera, Pau, France

Figure 1. A locally coherent event picked in the data volume. Described by a slope p_r, a receiver R and a two-way time T_{sr} determined in the common-shot gather and a slope p_s determined in common-receiver gather for the same shot S and two-way time T_{sr}.

locally coherent events are amenable to dense picking and hence high-resolution tomography. Indeed, the notion of locally coherent events opposes the conventional reflection tomography relying on exhaustive and subjective picking of laterally coherent reflection events. On the same line of thought, Billette *et al.* (1998) proposed stereotomography, a ray-based slope tomographic method relying on the semi-automatic picking of locally coherent events tied to scattering points in depth by a pair of ray segments (see Lambaré 2008 for a review). Each event is parametrized by its picked two-way traveltime and slopes (horizontal component of the slowness vector) at the source and receiver positions (Fig. 1), while the model space involves the scatterer coordinates (the starting points of the rays), several ray attributes (take-off angles and one-way traveltimes) and the velocity field. The inverse problem is implemented by building explicitly the sensitivity matrix and the resulting sparse tomographic system is solved with a linear conjugate-gradient method during each iteration of the velocity model update, this update being either performed in a linear or nonlinear way. Since the original formulation (Billette 1998), different variants emerged; for example, 3-D extension (Chalard *et al.* 2000), post-stack formulation (Lavaud *et al.* 2004), application in borehole settings (Gosselet *et al.* 2005), adaptation for anisotropic media (Nag *et al.* 2006; Barbosa *et al.* 2008), accounting for converted primary waves (Alerini *et al.* 2007) or wide-aperture data (Prieux *et al.* 2013), triangulated model parametrization (Yang *et al.* 2018), handling complex topography (Jin & Zhang 2018). All of the aforementioned variants follow the same framework of the classical formulation using ray tracing as a forward solver and explicitly building the sensitivity matrix for the inversion.

Recently, Tavakoli *et al.* (2017) and Tavakoli *et al.* (2019) developed an alternative formulation of slope tomography, referred to as adjoint slope tomography (AST), based on eikonal solvers (Fomel *et al.* 2009) and the adjoint-state method (Plessix 2006). The dissimilarity between AST and the original ray-based approach of Billette *et al.* (1998) goes beyond the manner of solving the forward problem (ray tracing versus finite-difference eikonal solver) in the sense that a more frugal parametrization of the data and model spaces was used in AST. However, a common feature of the two approaches is the joint update of the velocity model and the scatterer coordinates, which is amenable to the well-known ill-famed velocity–position coupling.

Other closely related methods such as the so-called Normal-Incidence Point (NIP) wave tomography based upon kinematic attributes picked in common-reflection surface (CRS) stack even put to use the second-order derivatives of traveltimes (Gelchinsky *et al.* 1999; Duveneck 2004). Dummong *et al.* (2008) conducted a direct comparison between stereotomography and NIP wave tomography and concluded that stereotomography provides velocity models of higher lateral resolution. This results from the different approximation of the traveltime curves used by the two methods: a local first-order local approximation in stereotomography amenable to the description of complex traveltime curves versus a hyperbolic approximation in NIP wave tomography that reduces the applicability of the method to moderate laterally inhomogeneous media. Bauer *et al.* (2017) revisited the now so-called wavefront tomography by exploiting the diffractions in order to tackle the aforementioned disadvantages.

Velocity model building was also recast in the framework of wave-equation tomography, a method that was proven superior to ray-based tomography in complex media but as might be expected is more computationally demanding (Luo & Schuster 1991). Clément *et al.* (2001) revamped migration-based traveltime inversion (MBTT; Plessix *et al.* 1999) using a waveform-based modelling engine. Furthermore, MBTT has been an inspiration for the development of the so-called reflection waveform inversion (RWI; Xu *et al.* 2012; Brossier *et al.* 2015; Wu & Alkhalifah 2015) or even its extensions, as for instance the incorporation of diving waves (Zhou *et al.* 2015). In a similar manner to reflection

traveltime tomography, RWI updates the velocity macromodel along the transmitted (or forward scattered) wavepaths (by opposition to ray paths) connecting the reflectivity inferred from least-squares depth migration to the sources and receivers at the surface. The velocity macromodel and the reflectivity can be updated jointly or in an alternating mode (Xu *et al.* 2012). Whatever the chosen strategy, the issues of these waveform-based approaches are related to the computational burden of full-waveform modelling as well as their potential sensitivity to amplitude errors and cycle skipping (Brossier *et al.* 2015; Zhou *et al.* 2018).

As an alternative to tomography, the other widespread family of velocity macromodel building methods would group the different variants of migration-based velocity analysis (MVA; Al-Yahya 1989). Unlike tomography methods that minimize residuals of seismic attributes in the data domain, MVA methods minimize residual moveout of reflection events in the migrated domain. Many adaptations and misfit criteria have been proposed for MVA, for example, differential semblance optimization (DSO; Symes & Carazzone 1991; Chauris & Noble 2001), wave-equation MVA (Shen *et al.* 2003; Sava & Biondi 2004) or more recently inversion velocity analysis (Chauris & Cocher 2017; Li & Chauris 2018).

Interestingly, Chauris *et al.* (2002a) established the relationship between data-domain stereotomography as developed by Billette (1998) and image-domain MVA based on locally coherent events as DSO (Symes & Carazzone 1991), hence reconciling these two popular families of velocity model building methods in an unified framework. This relationship was established by recognizing that the position of a scattering point in a 2-D subsurface medium can be found by migration of two kinematic attributes defined in the acquisition domain (e.g. two-way traveltime and one slope). This kinematic migration amounts in fact to solve the two simple focusing equations provided in Chauris *et al.* (2002a, their eqs 10 and 11). The flexibility highlighted by this relationship is exploited in MVA to reduce the number of depth migration. In stereotomography, the needed kinematic attributes in the acquisition domain (traveltime and slope) became accessible by demigration of kinematic invariants (strutural dip, residual moveout and subsurface position) picked in the depth-migrated domain, the latter being advantageous for picking due to the higher signal to noise ratio (Nguyen *et al.* 2008). The notion of kinematic invariance in stereotomography (Guillaume *et al.* 2008; Montel *et al.* 2010) emphasizes on the invariance of the picked attributes after a kinematic migration/demigration process in any given background velocity model.

To introduce the main motivation of this study, we come back to the velocity–position coupling problem in reflection traveltime tomography that has been evoked at the beginning of this section. In practice, even though the goal of tomography is the recovery of wave speeds, reflection tomography intrinsically introduces the reflector positions as an additional parameter class. Updating the velocity field and the reflector coordinates simultaneously is obviously rational, yet not straightforward since the underlying inverse problem may be ill-posed due to the explicit coupling between the velocity distribution and the reflector positions. Many works addressed this issue: Stork & Clayton (1985) opted for an iterative strategy by alternating tomography for velocity estimation and depth migration for reflector positioning, hence mishandling the coupling effect. Stork & Clayton (1986) showed that such an alternating-direction strategy that breaks down a nonlinear problem into two subproblems does not manage efficiently the trade-off issue. The joint inversion of both the velocity structure and the reflector geometries seemed, therefore, inevitable in reflection tomography (Bishop *et al.* 1985; Farra & Madariaga 1988). Unsurprisingly, the same conclusions were drawn in stereotomography, the coupling between velocity and the ray attributes associated with a scattering point was proven to be significant (Billette 1998, pages 95–99). On the same line of thought, Pavlis & Booker (1980) and Spencer & Gubbins (1980) discussed the detrimental impact of similar relaxation strategies in the context of earthquake hypocentres relocation and velocity estimation. More recently, Valensi *et al.* (2017) revisited this coupling problem in the context of RWI and showed the impact of different possible optimization strategies. In the latter, a third strategy was proposed called the reflectivity/background consistent approach that was shown to be superior. The consistent strategy relies on a variable projection approach (Golub & Pereyra 2003) which explicitly ties together the velocity field and the reflectivity through physical constraints.

Similar to the variable projection strategy promoted by Valensi *et al.* (2017), the objective of this paper is to revisit the AST parametrization by making use of the focusing equations of Chauris *et al.* (2002a) as constraints to position the scatterers in depth and reduce once more the model parametrization of slope tomography. In practice, the scattering points are first positioned in the current velocity model through a kinematic migration of a part of the data attributes (in our case, the two-way traveltime and one of the two slopes). Then, the retrieved scattering positions are projected into the constrained objective function before updating the velocity field by inversion of the residuals of the remaining slope. This variable projection (or elimination) mitigates the velocity–position coupling by ensuring the consistency between the scatterer positions and the velocity model in the migration sense. By doing so, we end up with a monovariate optimization problem involving one data class. In consequence, no scaling in the data and model spaces is needed. Following Tavakoli *et al.* (2017) and Tavakoli *et al.* (2019), we implement this new formulation of AST with the reduced-spaced method of Lagrange multipliers reviewed by Plessix (2006).

In the first section, we review the parametrization of the classical stereotomography and the slope tomography based on eikonal solvers and the adjoint-state method (AST). Then, we develop our proposed formulation under the same framework as AST. An application to the synthetic Marmousi case validated by FWI and another on real data validated through a TTI Kirchhoff migration will be presented. The different applications show the improved convergence speed of our approach relatively to AST, a more stable behaviour of the inversion and a more straightforward implementation due to the lack of scaling in the data and model spaces. In the final section, we discuss the impact of such a reduction in the model space and its implication on the coupling between velocity and scattering positions in comparison with other inversion strategies.

2 ON THE PARAMETRIZATION OF SLOPE TOMOGRAPHY

2.1 Classical stereotomography

In classical stereotomography (Billette 1998; Lambaré 2008), a locally coherent event n in the pre-stack data volume is defined by the source and receiver positions $(\mathbf{x}_s, \mathbf{x}_r)$, two-way traveltime T and horizontal component of the slowness vector at source and receiver positions (p_s, p_r) (Fig. 1). An all-inclusive parametrization was chosen in this ray-based approach, where the source and receiver positions were introduced as objective measures in order to absorb experimental errors and relax a boundary condition for the ray tracing problem. The corresponding optimization parameters can be subdivided into two categories. The first is related to the ray segments connecting the diffraction point to source and receiver positions: this includes scatterer coordinates $(\mathbf{x})$, take-off angles (ϕ_s, ϕ_r) and one-way traveltimes (T_s, T_r). The other involves wave speeds parametrized by B-spline coefficients c. This leads to the following definition of the data space $\mathbf{d}_{cs}$ and model space $\mathbf{m}_{cs}$ of classical stereotomography (cs)

$$\mathbf{d_{cs}} = [(\mathbf{x}_s, \mathbf{x}_r, T_{s,r}, p_s, p_r)_{n_{s,r}}]_{s=1}^{N_s}|_{r=1}^{N_r^s}|_{n_{s,r}=1}^{N_{n_{s,r}}} \tag{1a}$$

$$\mathbf{m_{cs}} = [(\mathbf{x}, \phi_s, \phi_r, T_s, T_r)_{n_{s,r}}]_{s=1}^{N_s}|_{r=1}^{N_r^s}|_{n_{s,r}=1}^{N_{n_{s,r}}}, [\mathbf{c}_m]_{m=1}^M. \tag{1b}$$

Following the notation of Tavakoli *et al.* (2017) in the above equations, M is the number of cubic B-spline nodes, N_s the number of shots, N_r^s the number of receivers for the shot s, $N_{n_{s,r}}$ the number of events for a source/receiver pair (s, r), $\mathbf{x}_{n_{s,r}}$ the coordinates of the nth scatterer tied to the source–receiver pair (s, r). Note that, with this parametrization, a common-reflection point in the subsurface sampled by n source–receiver offsets will be processed as n independent optimization parameters during slope tomography.

A Fréchet derivative matrix is explicitly constructed through paraxial ray tracing (Farra & Madariaga 1987) in order to solve the Newton-based local optimization scheme (Nocedal & Wright 2006). The optimization workflow includes three distinct steps: first the scattering positions $\mathbf{x}_{n_{s,r}}$ are initialized independently of the initial background velocity model using simplistic geometrical assumptions (refer to the Appendix A of Billette *et al.* 2003 for analytical expressions), followed by a so-called localization step where the ray attributes are solely updated, while keeping the starting velocity model fixed. In the final step, the totality of the parameters are updated jointly through an iterative nonlinear inversion. During each nonlinear iteration, the tomographic system is solved iteratively with a linear conjugate-gradient LSQR algorithm (Paige & Saunders 1982).

2.2 Adjoint slope tomography (AST)

In AST (Tavakoli *et al.* 2017, 2019), the forward problem computes traveltime maps with eikonal solvers using $\mathbf{x}_s$ and $\mathbf{x}_r$ as injection points, hence discarding $(\mathbf{x}_s, \mathbf{x}_r)$ and $(\phi_s, \phi_r, T_s, T_r)$ from the data and parameter spaces (eqs 1a and 1b), leading to the more compact data and model spaces, respectively

$$\mathbf{d_{ast}} = [(T_{s,r}, p_s, p_r)_{n_{s,r}}]_{s=1}^{N_s}|_{r=1}^{N_r^s}|_{n_{s,r}=1}^{N_{n_{s,r}}} \tag{2a}$$

$$\mathbf{m_{ast}} = [\mathbf{x}_{n_{s,r}}]_{s=1}^{N_s}|_{r=1}^{N_r^s}|_{n_{s,r}=1}^{N_{n_{s,r}}}, [\mathbf{c}_m]_{m=1}^M. \tag{2b}$$

Due to the use of eikonal solvers and the reciprocity principle, the problem complexity becomes proportional to the number of non-redundant source and receiver positions whereas it scales to the number of scattering points in the ray-based approaches. The inverse problem relies on the same steps as the classical approach but the gradient of the objective function is calculated with the matrix-free adjoint-state method (Plessix 2006).

2.3 Parsimonious adjoint slope tomography (PAST)

The data and model spaces of AST can be further reduced via the connection between stereotomographic attributes in the acquisition (data) domain and the pre-stack depth-migrated domain (Chauris *et al.* 2002a). The migration of kinematic invariants during the iterations of the slope tomography is implemented by solving the two following common-shot focusing equations (Chauris *et al.* 2002a)

$$T_{n_{s,r}}^* = t_s(\mathbf{x}_{n_{s,r}}) + t_r(\mathbf{x}_{n_{s,r}})$$
$$p_{r,n_{s,r}}^* = p_{r,n_{s,r}}, \tag{3}$$

which we use as constraints in PAST. In eq. (3), $T_{n_{s,r}}^*$ and $p_{r,n_{s,r}}^*$ denote the picked two-way traveltime and receiver slope associated with the event $n_{s,r}$, while $t_s(\mathbf{x}_{n_{s,r}})$ and $t_r(\mathbf{x}_{n_{s,r}})$ denote the modelled one-way traveltimes along the paths connecting the source s and the receiver r to the scattering point position $\mathbf{x}_{n_{s,r}}$, respectively. The traveltimes are extracted from the traveltime maps $t_s(\mathbf{x})$ and $t_r(\mathbf{x})$ computed in the current velocity model with the eikonal solver using shot s and receiver r as injection points (Fig. 2). Employing the presented parametrization,
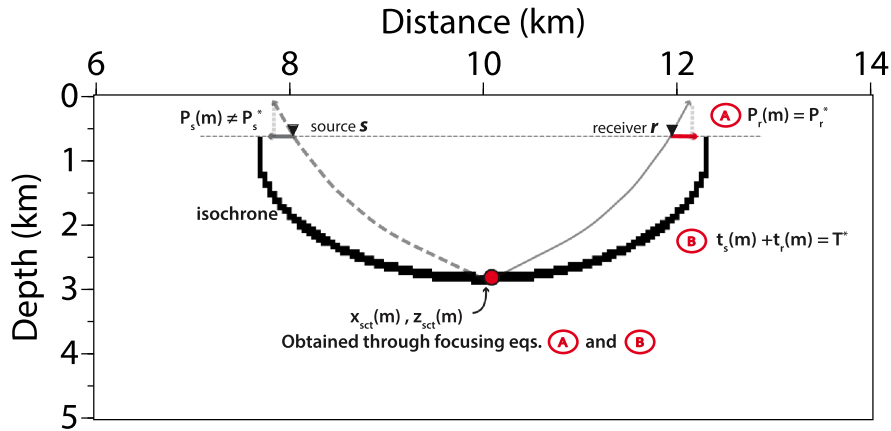
Figure 2. Focusing in depth a locally coherent event in the depth-migrated domain through the focusing equations of Chauris *et al.* (2002a).

the optimization boils down to invert for only one data class, here the source slope $p_{s,n_{s,r}}$ (eq. 4a) and reconstruct a sole parameter class in isotropic cases, namely wave speeds (eq. 4b)

$$\mathbf{d_{past}} = [p_{s,n_{s,r}}]_{s=1}^{N_s} |_{n_{s,r}=1}^{N_{n_{s,r}}} \tag{4a}$$

$$\mathbf{m_{past}} = [\mathbf{c}_m]_{m=1}^{M}. \tag{4b}$$

In the following section, we review in details the implementation of this data-space and model-space reduction in the framework of AST.

We would like to note that the idea of a kinematic migration by means of the focusing equations is well elaborated in the migration context. Indeed, the two-way traveltime and one slope are sufficient to define a locally coherent event in non-complex media (absence of triplications). The notion has been established in a migration context (Xu *et al.* 2001) based on the traveltime injectivity condition (TIC; ten Kroode *et al.* 1994).

3 METHOD

According to the proposed definition of the parsimonious data and model spaces (eq. 4), we aim to solve the following minimization problem

$$\min_{\mathbf{m}} J(\mathbf{m}) = \min_{\mathbf{m}} \frac{1}{2\sigma_{p_s}^{2}} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r^s} \sum_{n_{s,r}=1}^{N_{n_{s,r}}} \|(p_{s,n_{s,r}}(\mathbf{m}) - p_{s,n_{s,r}}^{*})\|^{2}, \tag{5}$$

where $p_{s,n_{s,r}}(\mathbf{m})$ denotes the predicted slope at the source and $\sigma_{p_s}^{2}$ are elements of a diagonal covariance matrix (Tarantola 1987). In practice, the inverse of this matrix will be used as a weighting operator.

Minimizing the objective function $J(\mathbf{m})$ with respect to $\mathbf{m}$ is a typical nonlinear tomographic problem that will be solved iteratively using a Newton-based local optimization scheme (Nocedal & Wright 2006). The predicted slopes $p_{s,n_{s,r}}$ depend on the model parameters $\mathbf{m}$ through a nonlinear forward problem operator $\mathbf{F}$, which gathers a series of physical (state) equations revolving around the calculation of traveltimes, the horizontal component of the source and receiver slowness vectors and the coordinates of the scatterers $\mathbf{x}_{n_{s,r}}$. Since $p_{s,n_{s,r}}$ has not a simple closed form expression with respect to $\mathbf{m}$, we enforce these physical constraints through a Lagrangian function and solve the constrained optimization problem with a reduced-space variable projection method generally referred to as the adjoint-state method (Haber *et al.* 2000; Plessix 2006). The Lagrangian function is written in compact form as

$$\mathcal{L}(\mathbf{m}, \mathbf{u}, \bar{\mathbf{u}}) = h(\mathbf{u}) - \langle \bar{\mathbf{u}} | \mathbf{F}(\mathbf{u}, \mathbf{m}) \rangle, \tag{6}$$

where $\langle . | . \rangle$ denotes the inner product, $\mathbf{u}$ gathers the state variables, $\bar{\mathbf{u}}$ the adjoint-state variables (or Lagrange multipliers) and $h(\mathbf{u}^*) = J(\mathbf{m})$ where $\mathbf{u}^*$ stands for a realization of the physical constraints.

3.1 State equations

We now review the different state equations gathered in $\mathbf{F}$.

We recall first the focusing equations presented by Chauris *et al.* (2002a) for the common-shot case

$$T_{s,r,n_{s,r}} = T_{s,r,n_{s,r}}^{*} \qquad p_{r,n_{sr}} = p_{r,n_{sr}}^{*}. \tag{7}$$

We invite the reader to refer to appendix A for a brief description on the practical manner of solving the focusing equations. We infer the predicted $T_{s,r,n_{s,r}}$ and $p_{r,n_{sr}}$ from traveltime maps $(t_s(\mathbf{x}), t_r(\mathbf{x}))$ computed with the fast sweeping method and a finite-difference factored eikonal

solver using the source and receiver positions as injection points (Fomel *et al.* 2009; Tavakoli *et al.* 2015).

$$H(\mathbf{x}, \nabla t_s(\mathbf{x})) = 0 \quad \text{with} \quad t_s(\mathbf{x}_s) = 0, \tag{8}$$

$$H(\mathbf{x}, \nabla t_r(\mathbf{x})) = 0 \quad \text{with} \quad t_r(\mathbf{x}_r) = 0. \tag{9}$$

In the above equations, we impose a Dirichlet boundary condition by zeroing the traveltimes at the source and receiver positions. The operator H stands for the Hamiltonian representation of the eikonal equation in tilted transversely isotropic (TTI) media (Alkhalifah 1998; Waheed *et al.* 2014). Its coefficients embed the model parameters we seek to update. We also introduce a sampling operator $Q_{n_{s,r}}$ implemented with a Kaiser-windowed sinc function (Hicks 2002) for the sake of traveltime extraction at any position in the traveltime maps $t_s(\mathbf{x})$ and $t_r(\mathbf{x})$.

From this, the two-way traveltimes $T_{s,r,n_{s,r}}$ are obtained straightforwardly by summing the traveltimes of t_s and t_r at the scatterer position $\mathbf{x}_{n_{s,r}}$

$$T_{s,r,n_{s,r}} = t_s(\mathbf{x}_{n_{s,r}}) + t_r(\mathbf{x}_{n_{s,r}}) = Q_{n_{s,r}} t_s(\mathbf{x}) + Q_{n_{s,r}} t_r(\mathbf{x}). \tag{10}$$

We estimate the receiver slopes in a finite-difference sense following the approach of Tavakoli *et al.* (2017). Spatial reciprocity allows us to infer $p_{r,n_{s,r}}$ from the values of the traveltime maps t_{r-1} and t_{r+1} initiated at neighbouring receivers $r-1$ and $r+1$ and sampled at $\mathbf{x}_{n_{s,r}}$, that is, far away from the injection points:

$$p_{r,n_{s,r}} = \frac{\partial T_{s,r,n_{s,r}}}{\partial x_r} = \frac{\partial t_r(\mathbf{x}_{n_{s,r}})}{\partial x_r} \approx \frac{Q_{n_{s,r}} t_{r+1}(\mathbf{x}) - Q_{n_{s,r}} t_{r-1}(\mathbf{x})}{2\Delta r}. \tag{11}$$

Even though the finite difference approximation is efficient, Qian & Symes (2002) pointed the inaccuracy of such approaches. A more precise strategy would involve solving an additional eikonal-based partial differential equation tying the traveltime perturbation with respect to the source position (Alkhalifah & Fomel 2010).

Substituting the expressions of $T_{s,r,n_{s,r}}$ and $p_{r,n_{s,r}}$, eqs (10) and (11), in the focusing eq. (3), allows us to eliminate the state variables $T_{s,r,n_{s,r}}$ and $p_{r,n_{s,r}}$ from our optimization problem. This gives

$$T^*_{s,r,n_{s,r}} = Q_{n_{s,r}} t_s(\mathbf{x}) + Q_{n_{s,r}} t_r(\mathbf{x}), \quad p^*_{r,n_{sr}} = \frac{Q_{n_{s,r}} t_{r+1}(\mathbf{x}) - Q_{n_{s,r}} t_{r-1}(\mathbf{x})}{2\Delta r}. \tag{12}$$

The two above focusing equations are used to estimate the scatterer coordinates $\mathbf{x}_{n_{s,r}}$, which are embedded in the sampling operator $Q_{n_{s,r}}$.

Finally, the state equations satisfied by the source slopes rely on the same finite-difference approximation as that used for the receiver slopes (eq. 11)

$$p_{s,n_{s,r}} = \frac{\partial T_{s,r,n_{s,r}}}{\partial x_s} = \frac{\partial t_s(\mathbf{x}_{n_{s,r}})}{\partial x_s} \approx \frac{Q_{n_{s,r}} t_{s+1}(\mathbf{x}) - Q_{n_{s,r}} t_{s-1}(\mathbf{x})}{2\Delta s}. \tag{13}$$

Injecting the state equations, eqs (8), (9), (12) and (13), in the Lagrangian function, eq. (6), gives

$$
\begin{aligned}
\mathcal{L}(\mathbf{m}, \mathbf{u}, \bar{\mathbf{u}}) = h(\mathbf{u}) - \sum_{s=1}^{N_s} \sum_{r=1}^{N_r^s} \sum_{n_{s,r}=1}^{N_{n_{s,r}}} & \left[\xi_{s,n_{s,r}} \left(p_{s,n_{s,r}} - \frac{Q_{n_{s,r}}(t_{s+1}(\mathbf{x}) - t_{s-1}(\mathbf{x}))}{2\Delta s} \right) \right. \\
& + \xi_{r,n_{s,r}} \left(p^*_{r,n_{s,r}} - \frac{Q_{n_{s,r}}(t_{r+1}(\mathbf{x}) - t_{r-1}(\mathbf{x}))}{2\Delta r} \right) + \mu_{s,r,n_{s,r}} \left(T^*_{s,r,n_{s,r}} - Q_{n_{s,r}}(t_s(\mathbf{x}) + t_r(\mathbf{x})) \right) \Big] \\
& - \frac{1}{2} \sum_{s=1}^{N_s} \Big\langle \lambda_s(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_s(\mathbf{x})) \Big\rangle_\Omega - \frac{1}{2} \sum_{r=1}^{N_r} \Big\langle \lambda_r(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_r(\mathbf{x})) \Big\rangle_\Omega,
\end{aligned}
\tag{14}
$$

where Ω denotes the subsurface domain Ω, $\mathbf{u} = \left(t_s(\mathbf{x}), t_r(\mathbf{x}), \mathbf{x}_{n_{s,r}}, p_{s,n_{s,r}} \right)$ and $\bar{\mathbf{u}} = \left(\lambda_s, \lambda_r, \xi_{r,n_{s,r}}, \mu_{s,r,n_{s,r}}, \xi_{s,n_{s,r}} \right)$. We remind the reader that the functional 14 is dependent to the model parameters in $\mathbf{m}$ through the eikonal equation present in the last two terms. The first-order optimality conditions, that is, the so-called Karush–Kuhn–Tucker conditions, state that a local minimizer of the constrained optimization problem is reached at the saddle point of the Lagrangian function (Nocedal & Wright 2006), that is when the three following equations are satisfied:

$$\partial \mathcal{L}/\partial \mathbf{u} = 0, \quad \partial \mathcal{L}/\partial \bar{\mathbf{u}} = 0 \quad \text{and} \quad \partial \mathcal{L}/\partial \mathbf{m} = 0. \tag{15}$$

Instead of updating jointly $\mathbf{u}$, $\bar{\mathbf{u}}$ and $\mathbf{m}$ during iterations, we resort to the reduced-space adjoint-state method (Haber *et al.* 2000; Plessix 2006) where the first two optimality conditions are strictly satisfied at each iteration through two successive variable projections. That is, we first estimate the state variables by solving the state equations $\partial \mathcal{L}/\partial \bar{\mathbf{u}} = 0$. From the estimated state variables, we estimate the Lagrange multipliers by solving the adjoint-state equations $\partial \mathcal{L}/\partial \mathbf{u} = 0$. Then, we infer the gradient of J with respect to $\mathbf{m}$ from the realizations of the state and adjoint-state equations by noting that $\nabla_{\mathbf{m}} J(\mathbf{m}) = \nabla_{\mathbf{m}} \mathcal{L}(\mathbf{u}^*, \bar{\mathbf{u}}, \mathbf{m})$. Finally, from $\nabla_{\mathbf{m}} J(\mathbf{m})$, we update $\mathbf{m}$ with an l-BFGS quasi-Newton iteration. This process is iterated until the convergence point where the three optimality conditions, eq. (15), are jointly satisfied. At this stage, it is worth reminding the key difference between AST and PAST formulations. In AST, $T^*_{s,r,n_{s,r}}$ and $p^*_{r,n_{sr}}$ are processed as objective measures (i.e. experimental quantities whose residuals are minimized by the objective function) and $\mathbf{x}_{n_{s,r}}$ as optimization parameters (quantities updated
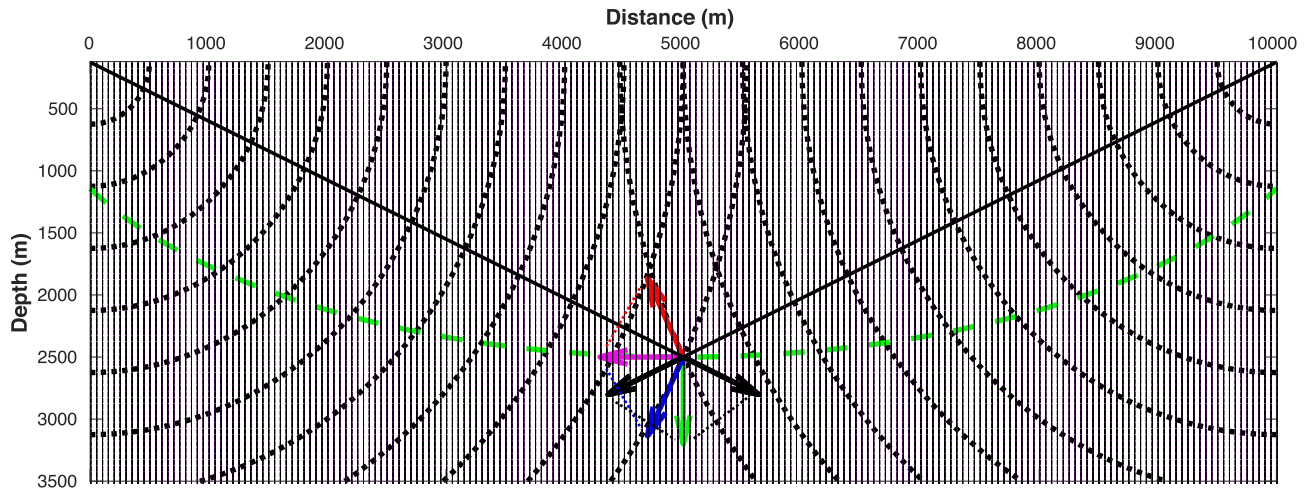
Figure 3. Graphical representation depicting the relationship between the rays (black lines), wavefronts (black dashed contours), slowness vectors (black arrows), gradient of slopes (red and blue arrows), the normal (green arrow) and the tangent (magenta arrow) vectors to the isochrone (green dashed curve). See the text for the interpretation of each vector.

by the optimization). In contrast, in PAST, $T^*_{s,r,n_{s,r}}$ and $p^*_{r,n_{sr}}$ are the right-hand sides of the two state equations satisfied by $\mathbf{x}_{n_{s,r}}$, which are now processed as state variables. In the framework of the reduced-space adjoint-state method, this indeed means that PAST strictly satisfies the focusing equations at each iteration, while AST introduces a relaxation of these equations by allowing two-way traveltimes and receiver slope residuals. To prevent any confusion, we emphasize on the fact that AST also relies on the reduced-space adjoint-state method, the difference with PAST being instead related to the parametrization of the data and model spaces. The relaxation of the focusing equations breaks down the kinematic consistency between the current velocity model and the position of the scatterers. In other words, the relaxation in AST is equivalent to solving the focusing equations (migration of the kinematic attributes) with velocities that differ from those of the current velocity model, while the demigration velocities used to compute the source slopes are the same in AST and PAST and correspond to the current velocity model. In this sense, AST enlarges the search space by relaxing two constraints, which might be beneficial to manage nonlinearities and absorb experimental errors related to inaccurate picking. However, the higher number of degrees of freedom this relaxation generates potentially increases the ill-posedness of the inversion by being more permissive in terms of velocity–position coupling.

3.2 Adjoint-state equations

We develop now the adjoint-state equations. The adjoint-state equation $\partial \mathcal{L}/\partial p_{s,n_{s,r}} = 0$ gives immediately that $\xi_{s,n_{s,r}}$ gather the scaled slope residuals

$$\xi_{s,n_{s,r}} = \frac{1}{\sigma^2_{p_s}}(p_{s,n_{s,r}} - p^*_{s,n_{s,r}}) = \frac{\Delta p_{s,n_{s,r}}}{\sigma^2_{p_s}}. \tag{16}$$

For each event, $\partial \mathcal{L}/\partial \mathbf{x}_{n_{s,r}} = 0$ gives the following 2×2 system of linear equations relating $\mu_{s,r,n_{s,r}}$, $\xi_{r,n_{s,r}}$ and $\xi_{s,n_{s,r}}$:

$$\mu_{s,r,n_{s,r}} \frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}}(\mathbf{t}_s + \mathbf{t}_r) + \frac{\xi_{s,n_{s,r}}}{2\Delta s} \frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}}(\mathbf{t}_{s+1} - \mathbf{t}_{s-1}) + \frac{\xi_{r,n_{s,r}}}{2\Delta r} \frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}}(\mathbf{t}_{r+1} - \mathbf{t}_{r-1}) = 0. \tag{17}$$

Unsurprisingly, the left-hand-side term of the system 17 has the same form as the gradient of the objective function with respect to $\mathbf{x}_{n_{s,r}}$ in AST (Tavakoli *et al.* 2017, their eq. 20) and gives clear insights on the relative role of two-way traveltimes and slopes in slope tomography. The weighted sum of the two slowness vectors at $\mathbf{x}_{n_{s,r}}$ (Fig. 3, black arrows) associated with the rays connecting $\mathbf{x}_s$ and $\mathbf{x}_r$ to $\mathbf{x}_{n_{s,r}}$, namely $\partial Q_{n_{s,r}}/\partial \mathbf{x}_{n_{s,r}}(\mathbf{t}_s + \mathbf{t}_r)$, is normal to the isochrone (Fig. 3, green arrow), while the gradient vectors $\partial Q_{n_{s,r}}/\partial \mathbf{x}_{n_{s,r}}(\mathbf{t}_{s+1} - \mathbf{t}_{s-1})$ and $\partial Q_{n_{s,r}}/\partial \mathbf{x}_{n_{s,r}}(\mathbf{t}_{r+1} - \mathbf{t}_{r-1})$ are normal to the rays connecting $\mathbf{x}_s$ and $\mathbf{x}_r$ to $\mathbf{x}_{n_{s,r}}$ (Fig. 3, red and blue arrows). Therefore, the first term (related to two-way traveltimes) controls the shifting of the scattering point $n_{s,r}$ perpendicularly to the isochrone, while the sum of the last two vectors (related to slopes) controls the drifting along the isochrone (Fig. 3, magenta arrow).

However, the Lagrange multipliers $\mu_{s,r,n_{s,r}}$ and $\xi_{r,n_{s,r}}$ differ in AST and PAST since they are tied to different state variables. Moving the terms depending on $\xi_{s,n_{s,r}}$ in the right-hand sides of the 2×2 system, eq. (17), and solving this system with Cramer's rule gives the closed form expression of $\mu_{s,r,n_{s,r}}$ and $\xi_{r,n_{s,r}}$ as function of $\xi_{s,n_{s,r}}$, that is, the source slope residuals (eq. 16),

$$\xi_{r,n_{s,r}} = -\xi_{s,n_{s,r}} \frac{\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{s,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|}{\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|} = -\mathcal{F}_r \xi_{s,n_{s,r}}, \tag{18}$$

$$\mu_{s,r,n_{s,r}} = -\xi_{s,n_{s,r}} \frac{\det \left| \frac{\partial(p_{s,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|}{\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|} = -\mathcal{F}_T \xi_{s,n_{s,r}}.$$ (19)

These expressions show that $\xi_{r,n_{s,r}}$ and $\mu_{s,r,n_{s,r}}$ are scaled versions of the source slope residuals. Injecting expressions of $\xi_{r,n_{s,r}}$ and $\mu_{s,r,n_{s,r}}$ in eq. (17) gives

$$\frac{\partial p_{s,n_{s,r}}}{\partial p_{r,n_{s,r}}} \frac{\partial p_{r,n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} + \frac{\partial p_{s,n_{s,r}}}{\partial T_{s,r,n_{s,r}}} \frac{\partial T_{s,r,n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} = \frac{\partial p_{s,n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}},$$ (20)

where

$$\frac{\partial p_{s,n_{s,r}}}{\partial p_{r,n_{s,r}}} = \frac{\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{s,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|}{\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|} \quad and \quad \frac{\partial p_{s,n_{s,r}}}{\partial T_{s,r,n_{s,r}}} = \frac{\det \left| \frac{\partial(p_{s,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|}{\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|}.$$ (21)

Indeed, eq. (20) is nothing else than the chain rule of derivatives. Remembering that the right-hand sides of the adjoint-state equations contains the partial derivative of $h(p_{s,n_{s,r}})$ with respect to the states $\mathbf{u}$ (Plessix 2006), the chain rule of derivatives highlights mathematically how the information carried out by p_r and $T_{s,r}$ on the subsurface velocities (left-hand side of eq. 20) are passed onto the optimization measure p_s (right-hand side of eq. 20) via the state variables $\mathbf{x}_{n_{s,r}}$, that is, the unknowns of the focusing equations whose right-hand sides are p_r^* and $T_{s,r}^*$ measurements. While $\xi_{r,n_{s,r}}$ and $\mu_{s,r,n_{s,r}}$ are the receiver slope and the two-way traveltime residuals in AST, they are now weighted versions of the source slope residuals, where the weights are the partial derivative of the source slope with respect to the two-way traveltime and the receiver slope.

Interestingly, eqs (18) and (19) also impose a reflection condition, namely $\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right| \neq 0$, which is similar to the imaging condition defined by Chauris *et al.* (2002a, their eq. 42). The term $\det \left| \frac{\partial(T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial(x_{n_{s,r}}, z_{n_{s,r}})} \right|$ represents the one-to-one mapping performed by means of the focusing equations between one locally coherent event in the common-shot gather parametrized by $T_{s,r,n_{s,r}}$ and $p_{r,n_{s,r}}$ and the position of a scatterer. The reflection condition is only violated when the normal vector to the isochrone and the orthogonal vector to the ray are parallel, that would correspond to the grazing incidence, that is, a scattering angle of $180°$. In the latter case, an ambiguity about the position of the scattering point along the isochrone would arise, hence validating our formulation in a reflection setting only. The reader is referred to Tavakoli *et al.* (2018) for the adaptation of adjoint slope-tomography to first arrivals where scatterers are naturally removed from the formalism.

Following Tavakoli *et al.* (2017), we obtain the adjoint-state equations satisfied by $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ by taking the derivative of the augmented functional 14 with respect to t_s and t_r. This gives

$$\begin{aligned} \left(\nabla \cdot (\lambda_s(\mathbf{x}) \mathcal{U}_s) \right)_\Omega = &\frac{1}{2\Delta s} \sum_{r=1}^{N_r^{s+1}} \sum_{n_{s+1,r}=1}^{N_{n_{s+1,r}}} Q_{n_{s+1,r}}^t \Delta p_{s+1,n_{s+1,r}} \\ &- \frac{1}{2\Delta s} \sum_{r=1}^{N_r^{s-1}} \sum_{n_{s-1,r}=1}^{N_{n_{s-1,r}}} Q_{n_{s-1,r}}^t \Delta p_{s-1,n_{s-1,r}} \\ &+ \sum_{r=1}^{N_r^s} \sum_{n_{s,r}=1}^{N_{n_{s,r}}} Q_{n_{s,r}}^t \mathcal{F}_T \Delta p_{s,n_{s,r}}, \end{aligned}$$ (22)

$$\begin{aligned} \left(\nabla \cdot (\lambda_r(\boldsymbol{x}) \mathcal{U}_r) \right)_\Omega = &-\frac{1}{2\Delta r} \sum_{s=1}^{N_s^{r+1}} \sum_{n_{s,r+1}=1}^{N_{n_{s,r+1}}} Q_{n_{s,r+1}}^t \mathcal{F}_{r+1} \Delta p_{s,n_{s,r+1}} \\ &+ \frac{1}{2\Delta r} \sum_{s=1}^{N_s^{r-1}} \sum_{n_{s,r-1}=1}^{N_{n_{s,r-1}}} Q_{n_{s,r-1}}^t \mathcal{F}_{r-1} \Delta p_{s,n_{s,r-1}} \\ &+ \sum_{s=1}^{N_s^r} \sum_{n_{s,r}=1}^{N_{n_{s,r}}} Q_{n_{s,r}}^t \mathcal{F}_T \Delta p_{s,n_{s,r}}. \end{aligned}$$ (23)

Adjoint fields $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ back-project the weighted sum of the source slope residuals along two ray tubes following the group velocity vectors $\mathcal{U}_s$ and $\mathcal{U}_r$ connecting $\mathbf{x}_{n_{s,r}}$ to $\mathbf{x}_s$ and $\mathbf{x}_r$, respectively (Fig. 4). The right-hand-side terms scale the amplitudes along the ray tubes based on the information carried out by traveltimes and slopes through the focusing weights $\mathcal{F}_T$ and $\mathcal{F}_r$ deduced from eqs (18) and (19). As the eikonal equations, the adjoint-state eqs (22) and (23) are solved with the fast sweeping method (Zhao 2005; Taillandier *et al.* 2009) using a conservative finite difference scheme as described by Tavakoli *et al.* (2019).
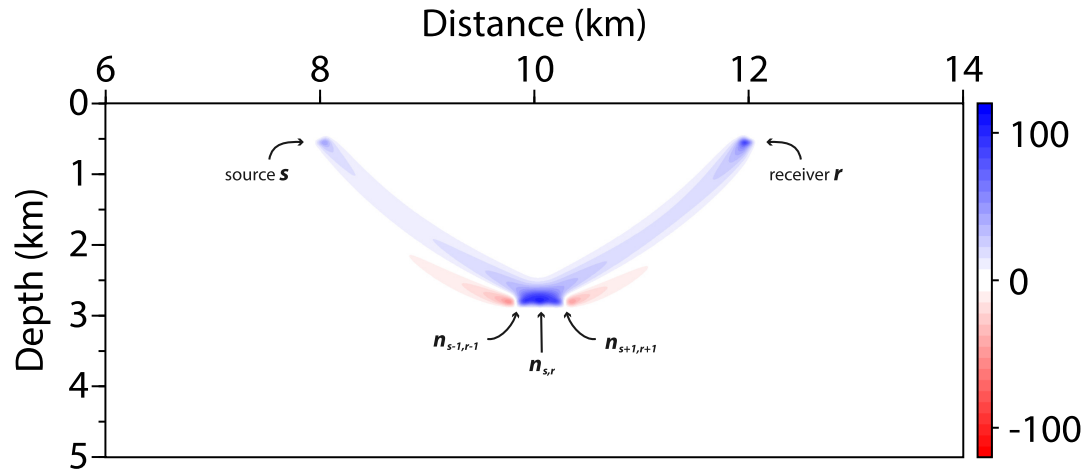
Figure 4. Sensitivity kernels λ_s and λ_r. The labels denote the source s, receiver r and three scattering points associated to the source/receiver per (s, r) and two virtual neighbouring pairs $(s - 1, r - 1)$ and $(s + 1, r + 1)$.

3.3 Gradient of the PAST objective function

From the adjoint-state variables, the gradient of the objective function $J(\mathbf{m})$ (eq. 14) with respect to the subsurface parameters is straightforwardly obtained by the weighted summation of the adjoint fields λ_s and λ_r

$$\nabla_{\mathbf{m}(\mathbf{x})} J = -\frac{1}{2} \left(\sum_{s=1}^{N_s} \frac{\partial H(\mathbf{x}, \nabla \mathbf{t}_s(\mathbf{x}))}{\partial \mathbf{m}(\mathbf{x})} \lambda_s(\mathbf{x}) + \sum_{r=1}^{N_r} \frac{\partial H(\mathbf{x}, \nabla \mathbf{t}_r(\mathbf{x}))}{\partial \mathbf{m}(\mathbf{x})} \lambda_r(\mathbf{x}) \right). \tag{24}$$

The weighting factors of the adjoint fields are simply the derivative of the forward operator $H(\mathbf{x}, \nabla t(\mathbf{x}))$ with respect to the model parameters and control how the gradients of J with respect to different parameter classes differ. Following the chain rule of derivatives, we then project the gradient on the cubic B-spline basis for the optimization (Tavakoli *et al.* 2017). We refer the reader to Tavakoli *et al.* (2019, Appendix B) for a first analysis of these multiparameter gradients in the case of TTI acoustic media.

To highlight the computational efficiency of the adjoint-state method relative to Fréchet derivative approaches and draw closer connections with the formulation of Chauris *et al.* (2002a), we review in the appendix B how the Fréchet derivatives of eikonal-based slope tomography could be implemented.

4 NUMERICAL EXAMPLES

In the following section, we analyse the performance of the reduced parametrization implemented in PAST in comparison to the joint inversion strategy employed in AST (Tavakoli *et al.* 2017). We benchmark the method on the well-known Marmousi model (Bourgeois *et al.* 1991) with further validation through time-domain FWI. We follow by a real application using high-end industrial BroadSeis data acquired in the Carnarvon Basin, north-west of Australia. The validation in the latter will be done through a TTI Kirchhoff migration.

4.1 Marmousi model

Recovering the long-wavelength component of the realistic Marmousi velocity model through tomographic approaches has been a challenge due to the strong lateral and vertical velocity changes (Chauris *et al.* 2002b; Billette *et al.* 2003). On top of that, the complexity of the structures hinders the reconstruction of the macromodel due to the ray theory shortcomings (Audebert *et al.* 1997). As in Tavakoli *et al.* (2017), we generate the data set by picking the positions and dips along the main reflectors of the true blocky model (Fig. 5a). In order to assess the inversion scheme, we calculate the data space using the same forward engine that will be used during the inversion. The demigration of the invariants is done in a smooth version of the Marmousi model, which is used as the target of the slope tomography (Fig. 5b). We generate this smooth model by applying a Gaussian isotropic filter of 100 m correlation length on the true model. The chosen 100 m correlation length ensures the validity of the single-arrival assumption in most part of the smooth model, while it preserves the kinematic properties of the true model (Operto *et al.* 2000). We however underline that in the targeted complex reservoir zone, some multipathing still occurs. The experimental setup mimics a towed-streamer acquisition consisting of 91 shots spaced 100 m apart and 134 receivers with an interdistance of 25 m. A total of 6708 picks were generated, as elaborated by Tavakoli *et al.* (2017), the events are associated to realistic specular reflection points.

We proceed with an inversion using the parsimonious PAST parametrization presented in the previous section. We remind the reader that there is no need for scaling in the data and model spaces since we solve a monovariate/monoparameter problem contrarily to AST. We note that the inversion can run using either the source or the receiver slopes as objective measures. In the Marmousi case, we opt for fitting the
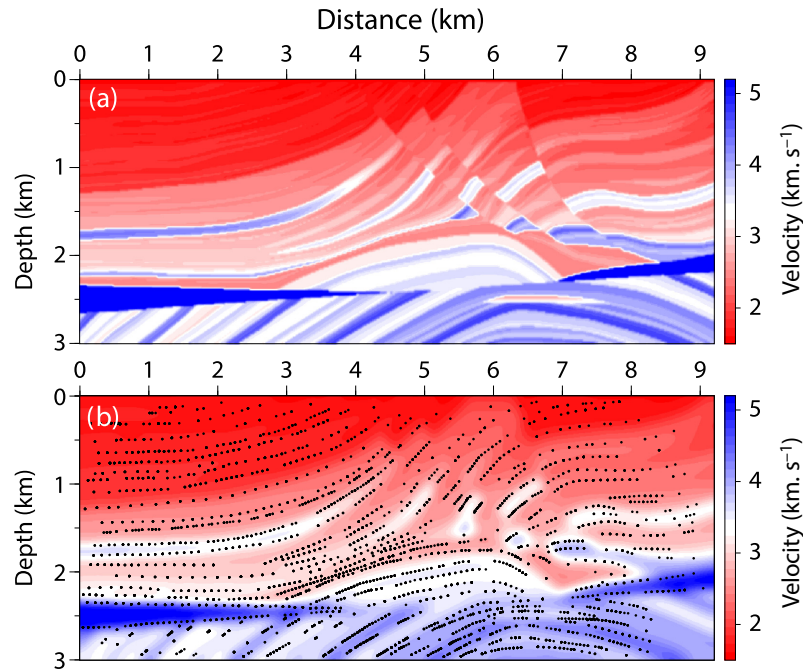
Figure 5. Marmousi example. (a) True blocky velocity model. (b) Picks of the Marmousi model reflectors. Picking was performed in the true velocity model shown in (a). Picks are superimposed on the smoothed velocity model that was used to generate the slope tomography measurements $p^*_{s,n_{sr}}$, $T^*_{s,r,n_{s,r}}$ and $p^*_{r,n_{sr}}$ by demigration.

receiver slopes due to a more favourable setup linked to the acquisition and the dip of the reflectors. The inversion is regularized by smoothing the gradient with a 200 m correlation length and by designing a multiscale reconstruction by successive B-spline refinements, in the same manner as Tavakoli *et al.* (2017) for the sake of an unbiased comparison.

During the inversion, we use the limited-memory Broyden-Fletcher-Goldfarb-Shannon (L-BFGS) algorithm (Byrd *et al.* 1995) implemented through the SEISCOPE optimization toolbox (Métivier & Brossier 2016) which also in turn manages the line search. Due to the use of a quasi-Newton optimization scheme, any scatterer that cannot be focused in the bounds of the computational domain is flagged as noise and kept out of the inversion until the memory dump occurs at the end of each scale.

As initial velocity model for PAST, we use the same homogeneous velocity model (2 km. s^{-1}) as that used by Tavakoli *et al.* (2017) to perform AST. This velocity model induces a quite incomplete subsurface illumination by the scatterers during the first iterations of PAST because the scatterers are positioned by the focusing equations (Fig. 6a). This trend is not seen during the early iterations of AST, where the initial scatterers are positioned by the initialization and localization steps described in Billette *et al.* (2003). In other words, while PAST initializes the scattering positions by migration to foster their kinematic consistency with the velocity model, the initialization of AST rather seeks to find the scattering positions that minimize the slope and traveltime residuals keeping the initial model fixed. Therefore, the initial incomplete subsurface coverage should not be seen as a disadvantage but rather as a consequence of the physical consistency between the background velocities and the scattering positions enforced by PAST. Of course, a different initial velocity model that spreads the scatterers in the totality of the computational domain would be more suitable for PAST.

During the early iterations of the inversion (Fig. 6b), the long-wavelength velocity distribution is retrieved, leading to a better scatterer coverage. The skeleton formed by the scatterers following the first three scales is analogous to the true structures especially up to 6 km distance and 2 km depth (Fig. 6b–d). In Fig. 6(e), the inverted tomographic model is presented along with a superimposition of the implicitly updated scattering point positions. We can clearly see that the aligned scatterers depict the overall structure of the reflectors found in the true model (Fig. 5a). In the later stages of the inversion, the updates are more concentrated on the complex part of the model around the reservoir (Fig. 6f,g). The final model (Fig. 6h) retrieved after a total of 195 iterations is quite satisfying, the reconstructed velocity is coherent with respect to the true model. Knowing that the superimposed scatterers in Fig. 6(h) are found by satisfying the focusing equations, we can draw a first conclusion that the background model is very suitable for migration (kinematics wise).

The misfit function evolution during the iterations exhibits a good behaviour (Fig. 7); the cost function value decreased by more than two orders of magnitude. The convergence is quite dramatic at the beginning of each scale due to the spline-grid refinement enforced when a line-search failure occurs. The increase in the cost function value seen at the start of some scales is due to the reintroduction of previously expelled events as explained earlier. We do not engage in a direct comparison with the misfit function behaviour seen in AST because of the various dissimilarities in its definition, the starting point of the inversion and the impact of scaling in the case of AST. In the following we focus on a comparison of the models and assess their validity as initial models for FWI.

In comparison with the result obtained by Tavakoli *et al.* (2017) after 374 iterations, the scattering points exhibit generally the same pattern (Fig. 8a) apart from the central reservoir region. The velocity logs shown in Fig. 9 prove that we were able to construct a smooth
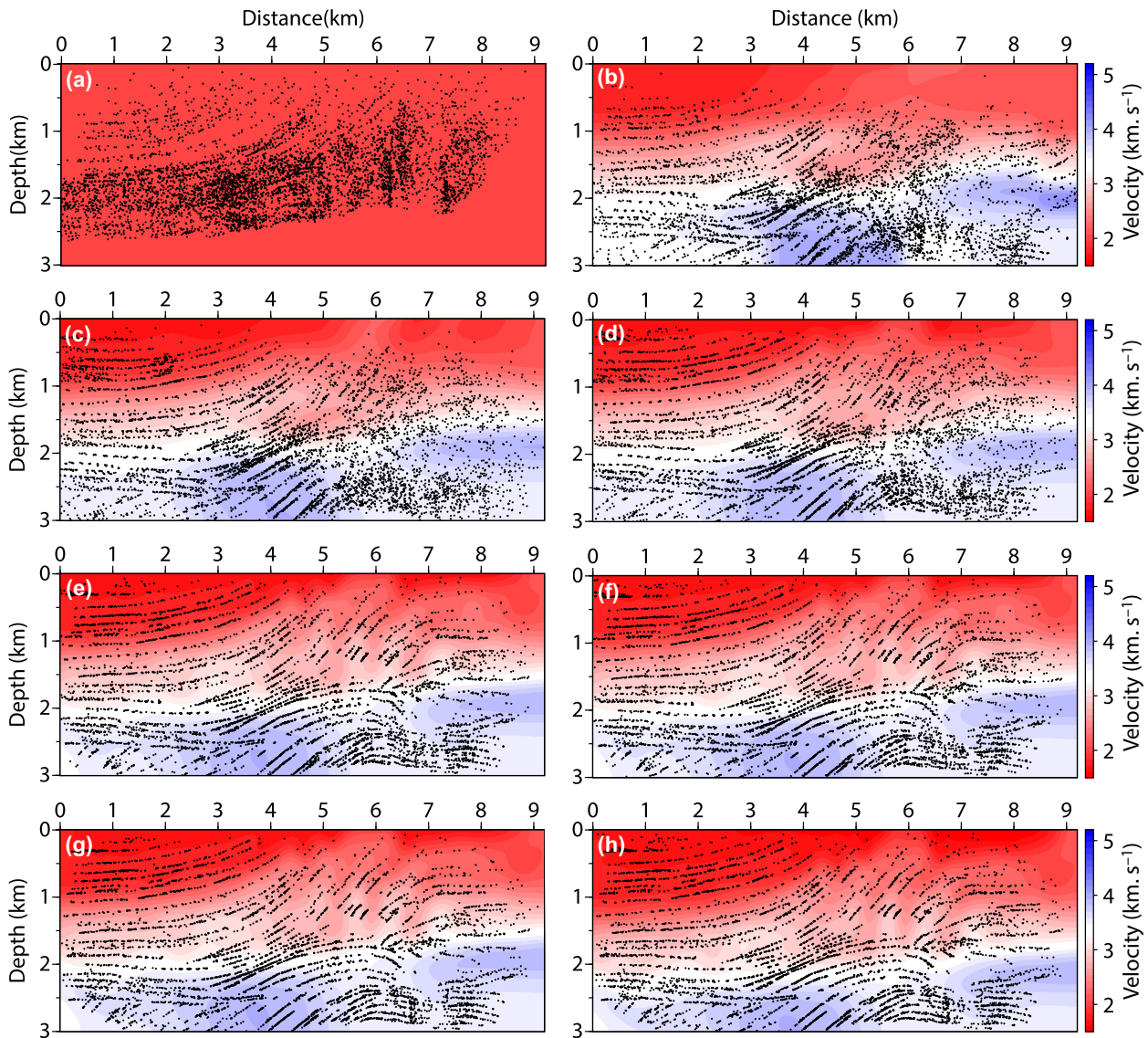
Figure 6. Marmousi example. Scatterers superimposed on velocity models inferred from PAST. Scatterer positions are found by solving the focusing equations. (a) Initial velocity model. (b–g) Velocity models inferred from PAST at the end of each multiscale step. (h) Final velocity model inferred from PAST.
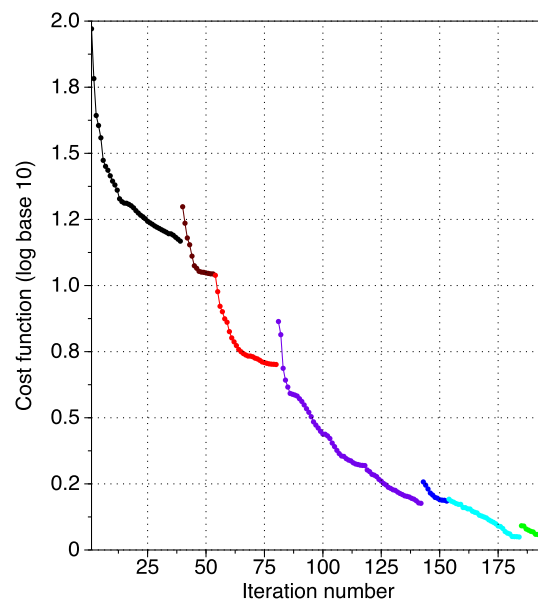

Figure 7. Marmousi example. Cost function value versus iteration number. The change in colours symbolizes a spline-grid refinement.
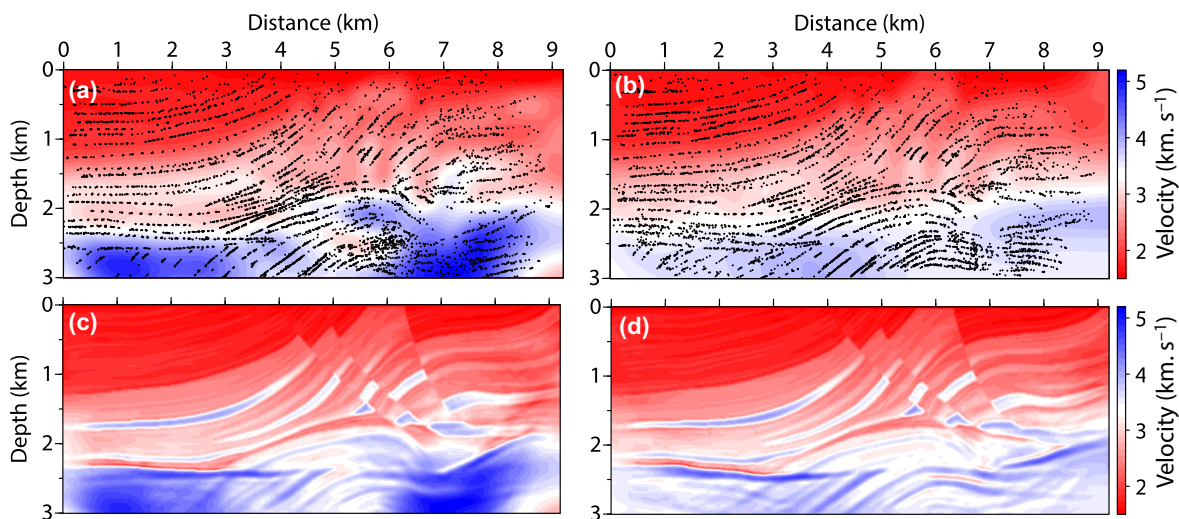
Figure 8. Marmousi example. (a) Velocity model inferred from 374 iterations of AST with superimposed scatterers (Tavakoli *et al.* 2017). (b) Velocity model inferred from 195 iterations of PAST with superimposed scatterers. In (a), scatterer coordinates were processed as optimization parameters in AST, while they were found by solving the focusing equations, eq. (12), in PAST. (c,d) Final FWI models obtained with initial models shown in (a) and (b).

version of the Marmousi model and that our inverted model follows a firmer trend than AST's (Fig. 9c). This trend shows that the inversion is more stable in some areas of the model, but we should state impartially that this could also be hindering the recovery of short wavelength structures.

We further assess the obtained velocity model as initial model for time-domain FWI. We consider a long-offset fixed-spread acquisition with sources and receivers covering the full surface to increase the sensitivity of FWI to cycle skipping. We invert successively seven data sets with increasing dominant frequencies [4, 6, 8, 10, 12, 14 and 16 Hz], a fixed lower bound of 2.5 Hz and a higher bound of 38 Hz. The FWI velocity model is in agreement with the true model down to a depth of 2 km (Fig. 8d). We can distinguish the better resolution around the reservoir in Fig. 8(d) especially the fact that the gas and oil cap is retrieved; this major discrepancy between both models is point on the log of Fig. 9(d) by green arrows. The vertical velocity logs (Fig. 9a–d) show inaccuracies in the deep part of the model below 2.5 km, as a consequence of the lack of illumination under the salt inclusions and the high-velocity contrast that is far from being recovered by tomography. We elaborate more on the reasons behind the convergence speed and stability differences in the discussion section.

4.2 Real TTI case study

We proceed with a real data application as validity assessment of the proposed approach. The concerned region is in the vicinity of the Carnarvon Basin, north-west of Australia. The area is well known for its hydrocarbon potential, especially the gas reservoir found at 3 to 4 km depth across the basin.

The data at hand is part of a previous exploration campaign. The 2-D profile is 57 km long. A total of 2479 shots were recorded by 648 receivers forming a 8250 m streamer. During the acquisition the streamer was submerged in a curvy manner at a depth up to 57.5 m; such settings induce a notch diversity that is well exploited for acquiring state-of-the-art data (Soubaras & Whiting 2011). The BroadSeis technology alleviates the low signal to noise ratio at low frequencies and higher frequencies accessibility. The advancement in acquisition design and equipment has been significant in the last decade; such developments unlock the potential of most seismic methods. Indeed, the broad-band nature of the new high-end seismic data is a step forward towards closing the ill-famed frequency gap (Claerbout 1985).

The processed data, local well measurements and the picked attributes for slope tomography were made available to us along with the estimated TTI parameters and most importantly a legacy velocity model built by 2-D nonlinear slope tomography (Fig. 10) which will only be used during the validation of the results.

In total, around two million events were picked in the depth migrated domain. During the demigration process a sort of redatuming is used to flatten the streamer line on the profile. Such processing causes inconsistencies in the retrieved source and receiver positions in the data domain. In other words, two picked scattering points associated with the same source–receiver pair will have shifted distinct source–receiver pairs in the data volume. This artificial increase in source and receiver positions is problematic from a computational point of view since the complexity of the problem under the adjoint framework is directly proportional to the number of non-redundant source/receiver positions. Tavakoli *et al.* (2019) implemented an aggregation scheme that would keep in the data space only scattering points associated with source/receivers positions shifted to a certain limited extent. This solution alleviates the computational burden. Only fifty thousand picks were selected by Tavakoli *et al.* (2019) for the AST inversion (Fig. 11a–c), as a consequence of the aggregation scheme and the elimination
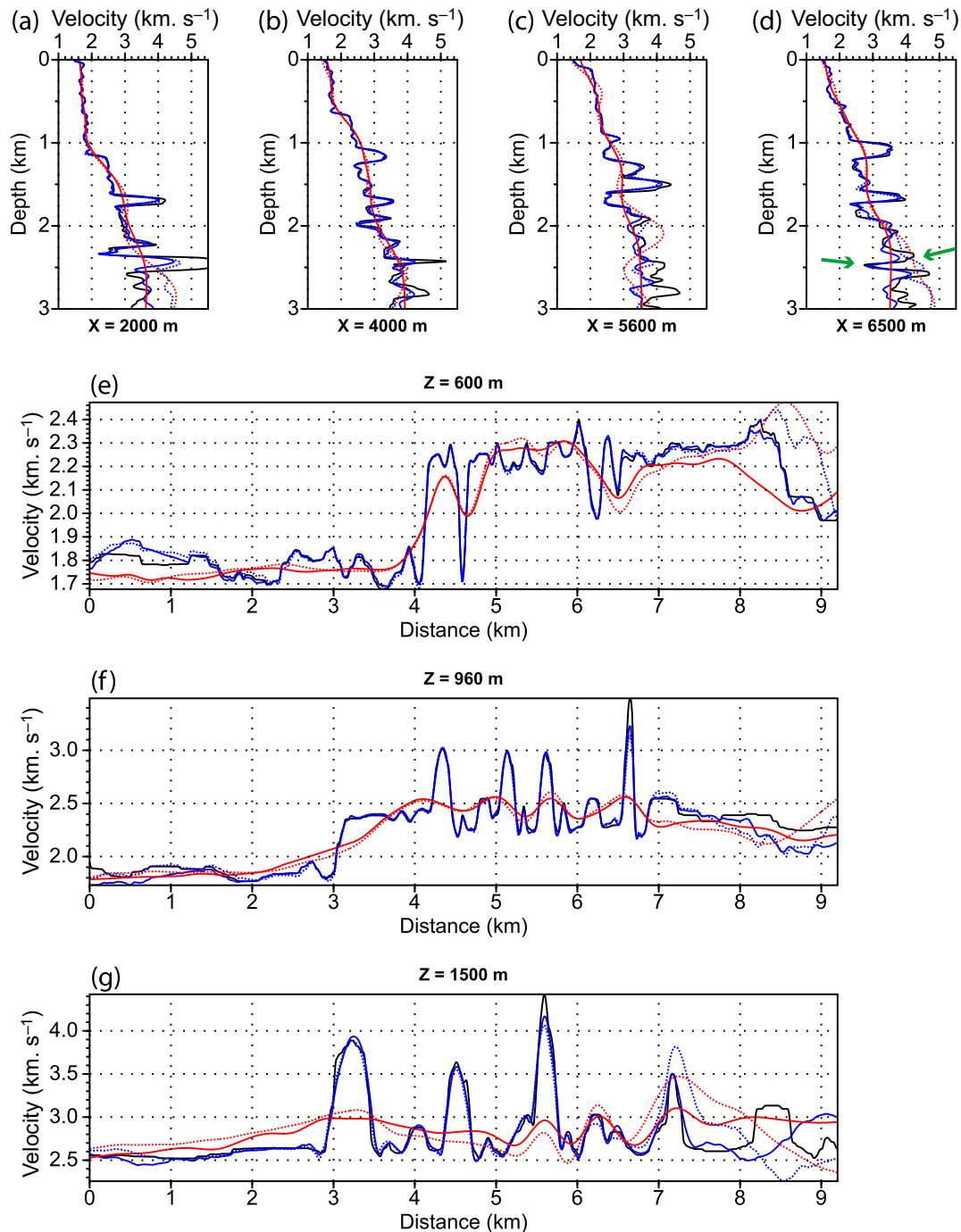
Figure 9. Marmousi example. Comparative velocity logs. The tomographic, FWI and true models are represented by red, blue and black lines, respectively. The dashed lines denote the results obtained by Tavakoli *et al.* (2017). The green arrows in (d) delineate the oil and gas cap low-velocity anomaly.

of suspected noisy picks. We note that this does not mean that only 2.5 percent of the data is reliable. We just expect that the pre-selected picks are sufficient for a good inversion across the totality of the model.

We start the inversion using a constant gradient model defined by $v = v_0 + a \times z_{bat}$, where z_{bat} is the depth from below the bathymetry line, $v_0 = 1500\,\mathrm{m\,s^{-1}}$ and $a = 0.5$ (Fig. 11d). This background velocity model is not optimal for our approach (Fig. 11d) but we prefer to use the same initial model as the one defined by Tavakoli *et al.* (2019) for the sake of consistency (Fig. 11a). We implement a six steps spline-grid refinement with horizontal and vertical node spacing decreasing from 2000 to 60 m and from 670 to 80 m, respectively. The difference between our experimental setup and the one employed by Tavakoli *et al.* 2019 (Fig. 11b,c) is the Gaussian filter correlation length (200 m instead of 500 m) and the fact that well logs data were not introduced as constraints during our inversion. The reason behind these discrepancies is to push the inversion to its limit and keep the well logs for a preliminary quality control.
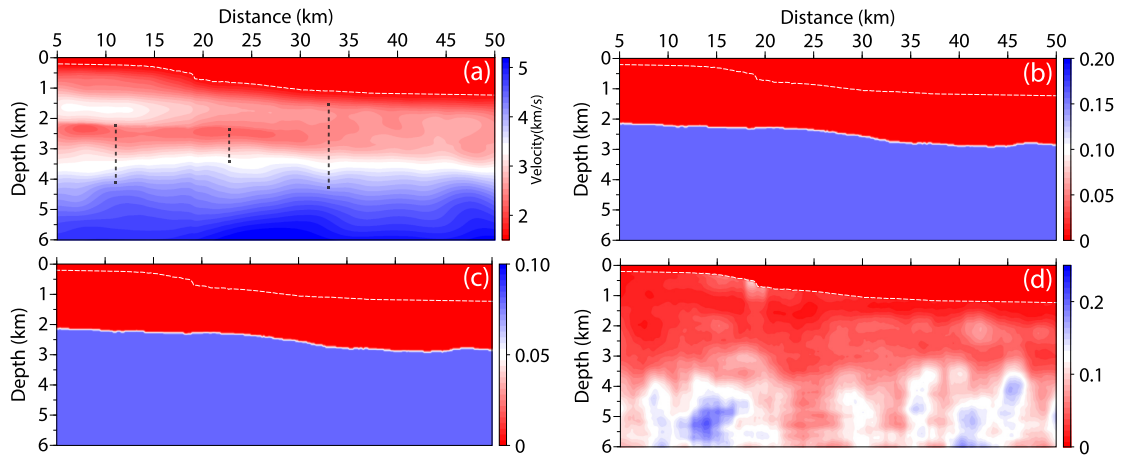
Figure 10. Real data application. (a) Legacy velocity model used as a reference. (b–d) Provided TTI anisotropy model parameters, Epsilon, Delta and Theta, respectively. The white dashed line denotes the bathymetry. The black dashed line in (a) denotes the available well data location.
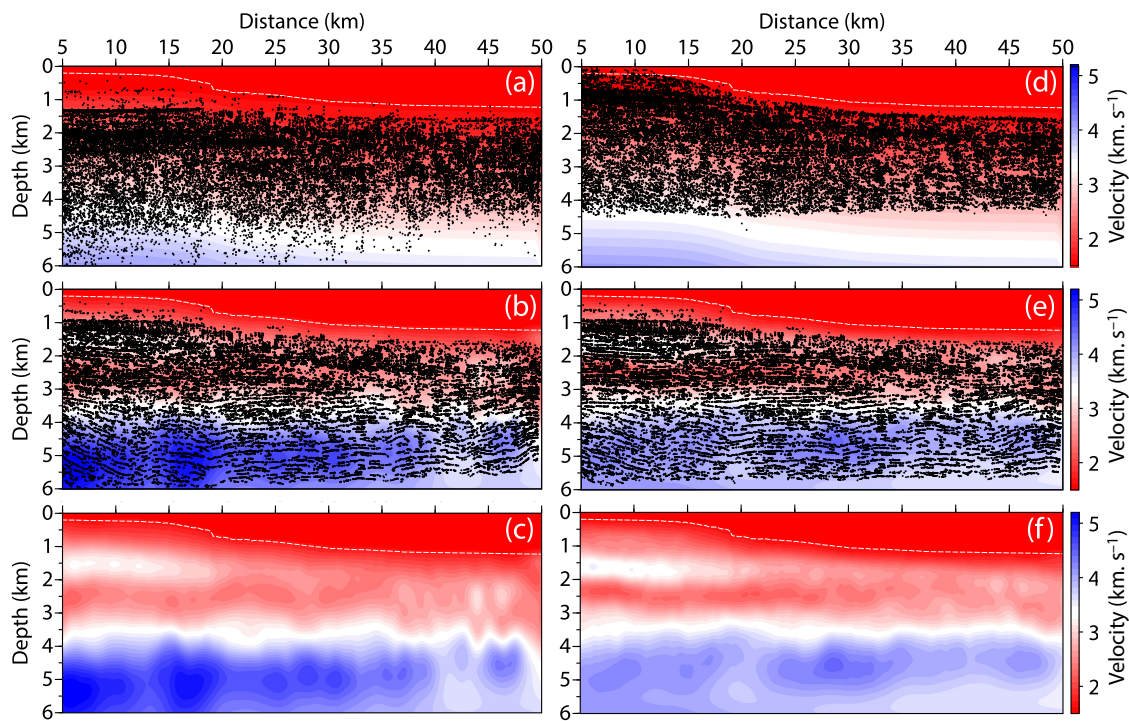

Figure 11. Real data application. (a) Constant gradient starting model with superimposed scatterers after initialization + localization in the case of AST. (b) Final velocity model obtained by AST superimposed by the final scatterer positions. (c) Final velocity model obtained. (d) Scatterer positions obtained by solving the focusing equations in the case of PAST using the same initial model. (e,f) same as (b,c) but in the case of PAST.

The final velocity model of PAST after 169 iterations is shown in Fig. 11(f). The velocity model was significantly updated in comparison to the starting model. At first glance, we can distinguish the laterally continuous low-velocity layer at 2.5 km depth and between 5 and 32 km distance. Another obvious feature is the building of high-velocity zones starting at 4 km depth even though very few picks correspond to offsets larger than 5 km. The ensemble of the final scattering positions depicts lateral continuously in a very coherent manner (Fig. 11e). The cost function value dropped of around 2.5 order of magnitudes with a very stable trend (Fig. 12).

The reconstructed velocity model is in agreement with the one obtained through AST (Fig. 11c) and the legacy model (Fig. 10a). Same features are found across all the models, especially between 35 and 50 km distance at 2 km same anomalies and shapes are recovered. We compare in Fig. 13 four vertical logs that are extracted from the legacy model, the AST model and the PAST model every 10 km. The velocity logs are mostly in agreement until 4 km depth where our retrieved model seem to have lower velocities. With further assessment, we will show that there is no underestimation of wave speeds in PAST. Another discrepancy between the three models is related to the low-velocity layer at 2.5 km depth and 10 km distance (Fig. 13a); the low-velocity zone reconstructed by AST is less sharp, which probably results from
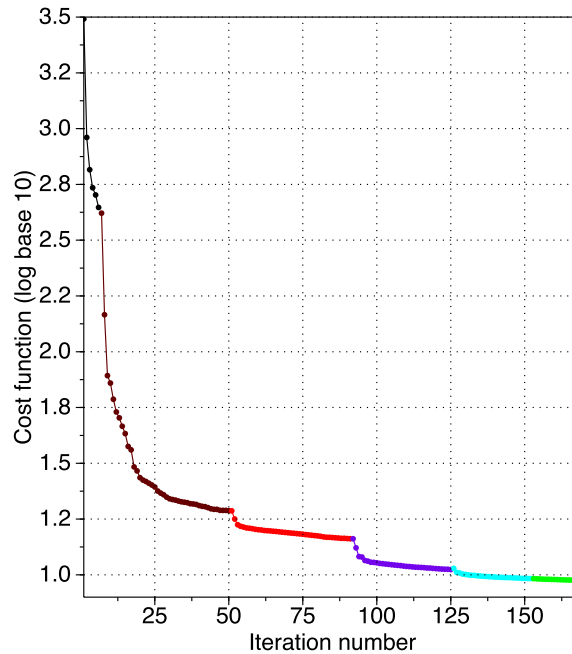
Figure 12. Real data application. Cost function value versus iteration number. The change in colours symbolizes a spline-grid refinement.
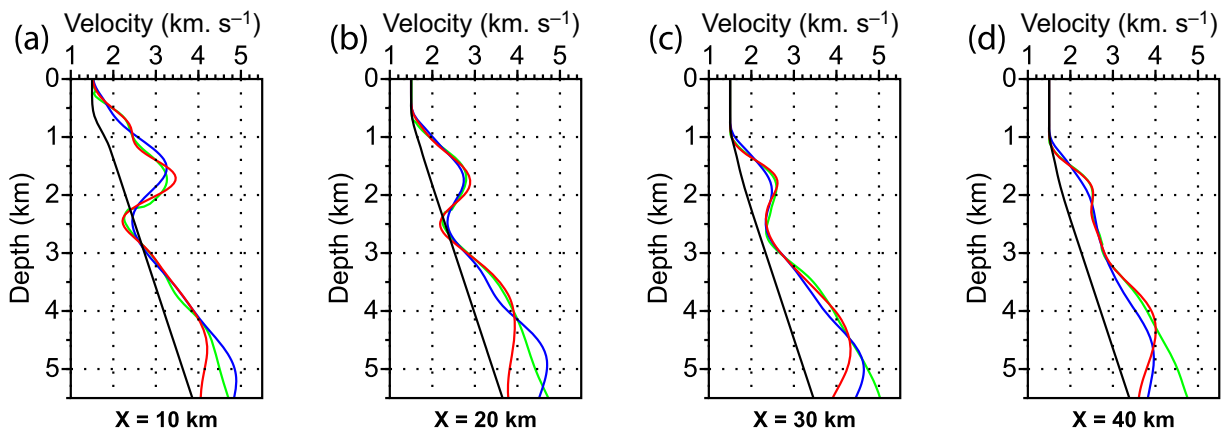

Figure 13. Real data application. Vertical velocity logs extracted from the starting model (black lines), the legacy reference model (green lines), the AST model (blue lines) and the PAST model (red lines). The logs are extracted at distances $x = 10$ km, 20 km, 30 km, 40 km.

the more aggressive regularization used during AST. The imprint of this regularization is illustrated by the smoother log reconstructed by AST relative to the two other inversions (Fig. 13).

As a first quality control, we check the validity of the PAST velocity model with respect to the available well log data. The logs show very matching trends and also show that at 4 km depth, there is no underestimation of the velocity (Fig. 14). As a more qualitative assessment, we compare TTI Kirchhoff pre-stack depth images. The TTI parameters shown in Fig. 10 were used for the migrations. The migrated images using the legacy model, the AST velocity model and the PAST velocity models as background models show the same structures (Fig. 15). The stratigraphy and the structural geology features are in accordance with the findings published in that region (Hocking 1988). The common image gathers (CIGs) confirm the similar accuracy of the three velocity models down to 4 km (Fig. 16). However, some events are flatter in the shallow part and more focused in the deep part of the CIGs computed with the PAST velocity model (Fig. 16, arrows and ellipse).

5 DISCUSSION

We have proposed a new formulation of eikonal-solver based adjoint slope tomography (PAST). This reformulation leads to a reduced (or parsimonious) parametrization of the data and model spaces, each of them involving one class of variable (one slope and wave speeds, respectively). This parsimonious parametrization is achieved by positioning the scattering positions through the migration of two kinematic invariants of the data space (two-way traveltime and one slope), this migration being implemented by solving the two focusing equations of Chauris *et al.* (2002a). This discards these two attributes from the data space and the scatterer coordinates from the model space, hence
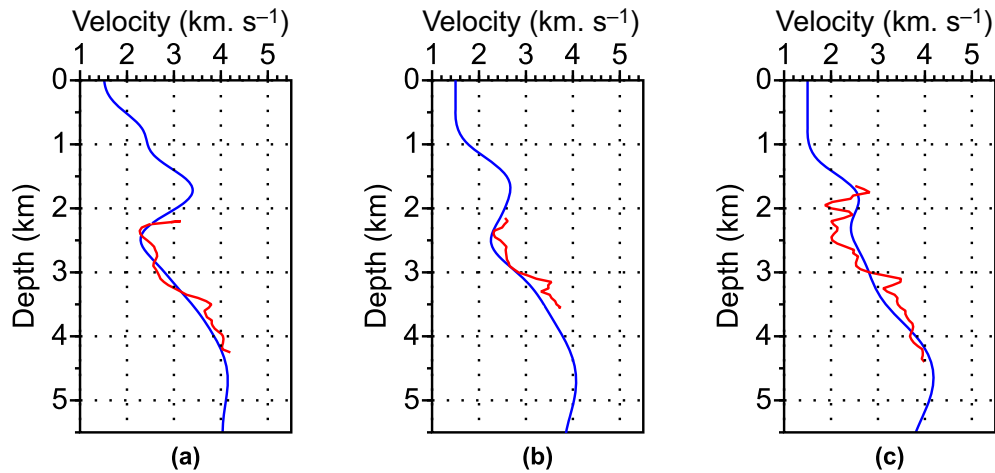
Figure 14. Real data application. Comparative velocity logs taken at 10.8 km (a), 22.6 km (b), 33.1 km (c). Inverted PAST velocity model and well data are, respectively, represented by blue and red lines.

leading to a better-posed inversion. In contrast, the recent AST developed by Tavakoli *et al.* (2017) and Tavakoli *et al.* (2019) introduces a relaxation of the focusing equations by processing the scattering positions as optimization parameters and the two kinematic invariants as optimization measures. This indeed implies that the positions of the scattering positions found by AST potentially differ from those obtained by PAST. Translating this statement in terms of velocities, the virtual velocities that would allow us to migrate the kinematic invariants at the positions of the scattering positions found by AST can differ from the background velocities of the current slope tomography iteration. In the following, the velocities that would give the positions of the scattering positions estimated by AST or PAST by migration of kinematic invariants are referred to as migration velocities, while the velocities used to solve the forward problem (computation of two-way traveltime and slopes) are referred to as demigration velocities. In PAST, the migration and demigration velocities are the same, while they can differ in AST.

Before discussing further the results of the two case studies presented in the previous section, we illustrate the concept of migration versus demigration velocities in the framework of slope tomography and its effect on the convergence path followed by AST and PAST with a toy example, inspired from the analysis of RWI performed by Valensi *et al.* (2017).

We use a homogeneous model in order to derive the solutions of the forward problem and the gradients analytically. The experimental setup consists of a true background velocity of 4.75 km. s^{-1} and a single source–receiver couple with an offset of 6 km tied to three random scattering positions in depth.

As a reminder, in the case of PAST scattering positions are focused in the initial model whereas in AST, the initial scattering positions are found by the initialization/localization step developed by Billette *et al.* (2003). Since the true model is homogeneous, the initialization step would position the scatterers at their true position with the equivalent medium being equal to the seeked velocity, that is due to the simplistic assumption that the initialization equations relie on. However, the localization step that aims at fitting the data while optimizing the position of the scatterers with respect to the initial background velocity would redistribute them across the model.

We note that following what is done in practice in AST, only few localization iterations were made in order to avoid an overfit of the data in the wrong velocity model and handicap the start of the joint inversion. A scaling on the data was applied in order to make the data dimensionless. In Fig. 17, we look at the optimization paths followed by AST and PAST. In order to examine the behaviour of the two approaches subject to the coupling management governed by the Hessian, we test three different optimization schemes (gradient descent, Gauss–Newton and BFGS) along with an inexact line search based on the strong Wolfe conditions (Nocedal & Wright 2006). We start the inversion with an initial model of 2.87 km. s^{-1}. In the case of PAST, the demigration and migration velocities are always consistent for all scattering points and the optimization path is following the diagonal straightly into the minimum. On the other hand in the case of AST, each scatterer is positioned by a different migration velocity. Although both AST and PAST converge towards the global minimizer using the employed optimization strategies, this test is enough to illustrate the slower convergence of AST related to PAST resulting from the relaxation of the focusing equations. As would be expected, the convergence path of AST is highly vulnerable to the accuracy of the Hessian's approximation. The Gauss–Newton scheme test reveals a competitive performance of AST with respect to PAST while on the other hand the gradient descent scheme exhibits a zigzag pattern commanded by the trade-off due to the absence of the Hessian. We remind the reader that in the previous applications presented in this paper, the Hessian was approximated through a quasi-Newton scheme that is comparable to the third scheme employed in this test (Fig. 17). The test reveals that during the early iterations of AST and in the absence of a good approximation of the Hessian, similar to the gradient descent test, the optimization is driven by velocity updates in order to compensate for the inconsistent positions delivered by the localization step. In subsequent iterations, due to the improved Hessian approximation the optimization path follows a consistent trajectory as the one exhibited by PAST. It is therefore reasonable to conclude that in complex cases where the inconsistency
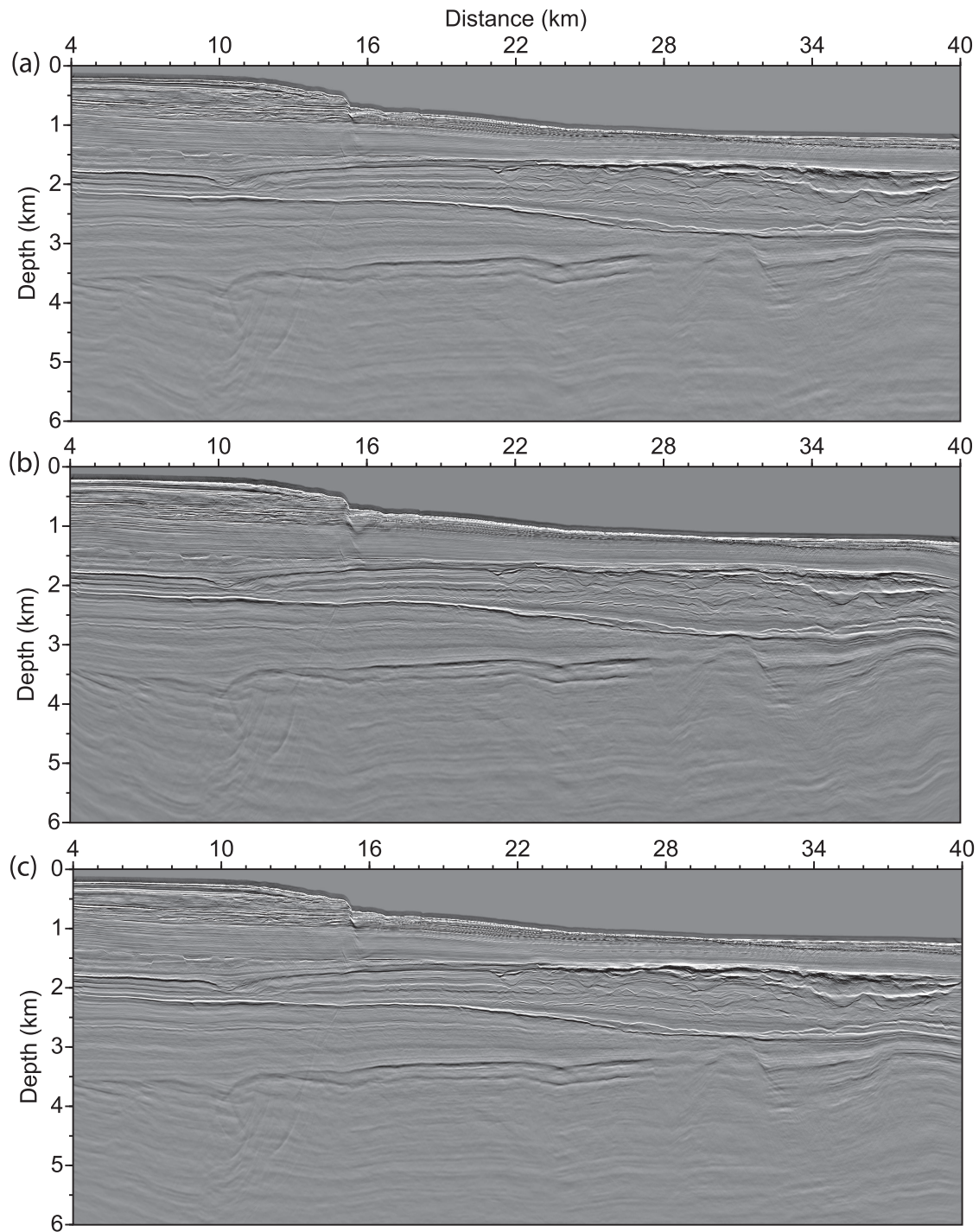
Figure 15. Real data application. Image obtained through TTI pre-stack depth Kirchhoff migration using as background model: (a) the reference legacy velocity model, (b) the AST velocity model and (c) the PAST velocity model.

between the migration/demigration velocity is aggravated and the coupling is not well managed as in this simple case, the performance of PAST would be even more superior with respect to AST.

This statement is now checked by comparing more closely the convergence histories of PAST and AST for the complex synthetic Marmousi benchmark. Fig. 18 shows the relative root mean square error (RMSE) of the velocity distribution (VRMSE) versus that of the scattering coordinates (SCRMSE) at each iteration. We use the same initial velocity model for AST and PAST and, hence both inversions start from the same VRMSE (vertical axis in Fig. 18). On the other hand, the initial SCRMSE is different (horizontal axis), since the localization step of AST estimates the initial scattering positions by minimizing in a least-squares sense the two slopes and two-way traveltime residuals using the initial velocity model as fixed background model, while PAST performs this positioning by solving the 2×2 focusing equations.
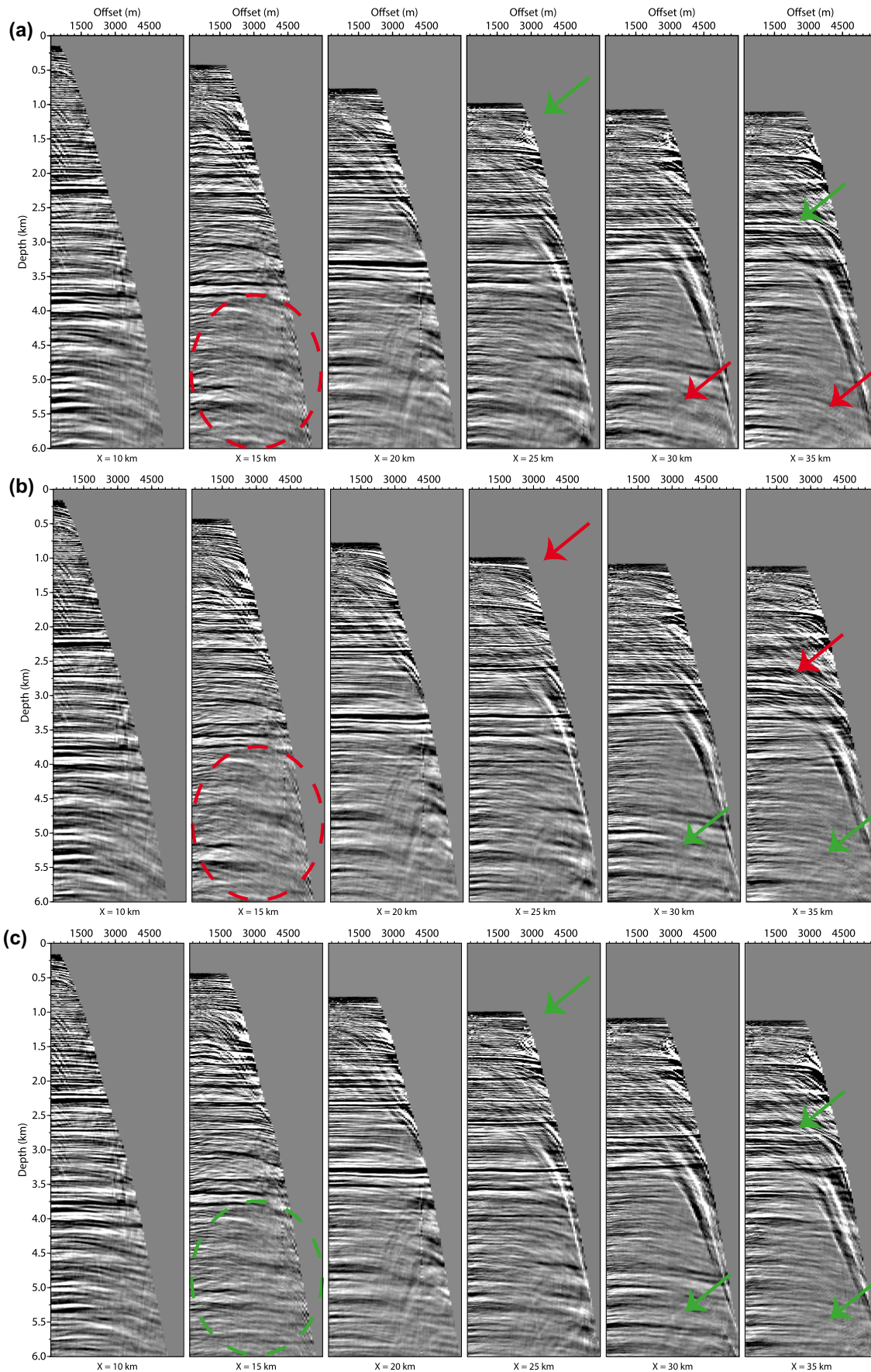
Figure 16. Real data application. Common image gathers collected at various positions corresponding to the migrated images shown in Fig. 15.
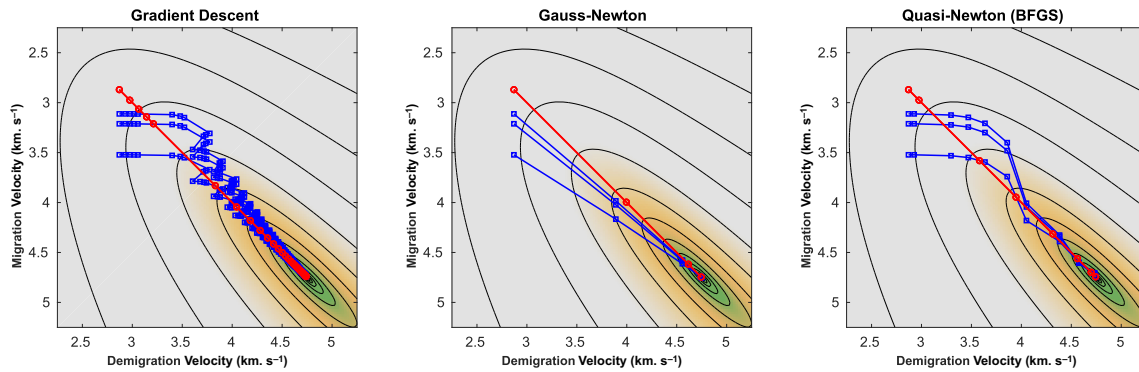
Figure 17. Toy test. Demigration velocity versus migration velocity. The red and blue lines denote the optimization paths taken by the proposed approach and AST (Tavakoli *et al.* 2017) respectively, while the red circles and blue squares denote the iterations. The black curves denote the cost function isovalues. Three different optimization strategies: On the left-hand side, Gradient Descent: 39 it. (PAST) – 362 it. (AST). In the centre, Gauss–Newton: 4 it. (PAST) – 5 it. (AST). On the right-hand side, Quasi-Newton (BFGS): 8 it. (PAST) – 11 it. (AST).
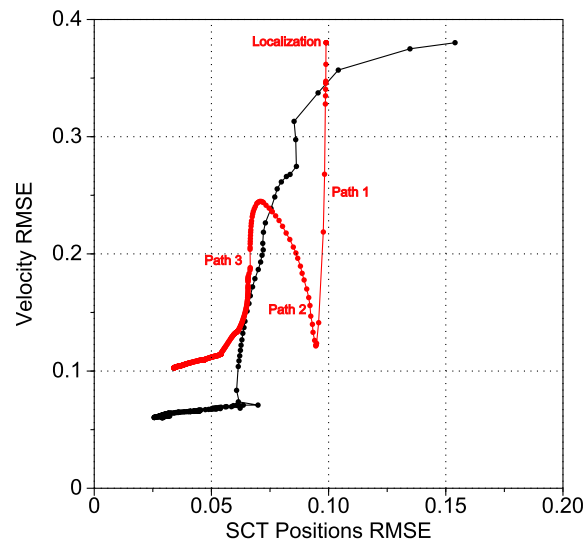

Figure 18. Marmousi example. Velocity versus scatterer position relative root mean square error evolution during the inversion. The red and black curves denote the inversion through AST (Tavakoli *et al.* 2017) and the proposed approach, respectively.

This indeed implies that the localization step is an overdetermined problem in AST since the scattering positions are estimated from the three data attributes, while the focusing equations show that two of these attributes are sufficient to unambiguously position the scatterers.

The path followed by AST (Fig. 18, red line) clearly shows that AST mostly updates velocity at the expense of scattering positions during the early iterations. This descent direction along the velocity axis results from the former localization step during which we best fit the data by updating the scattering positions keeping the initial velocity model fixed. Accordingly, AST is left after the localization step with residuals in which the signature of the scattering position errors has been removed, hence driving the subsequent inversion towards velocity updates. In other words, the localization step combined with the early iterations of AST lead to an alternating-direction inversion for scattering positions and wave speeds. Valensi *et al.* (2017) illustrated in the framework of RWI how these alternating-direction strategies can lead to slow convergence. During this velocity update, we see that the path followed by AST tends to move away from that followed by PAST (Fig. 18, black line). This reflects that the virtual migration velocities that would give the scattering positions estimated by AST become increasingly inconsistent with the demigration (background) velocities. Once the inversion reaches a point where the scattering positions retrieve a significant imprint in the data residuals, AST changes sharply its descent direction by updating the two parameter classes in a more balanced way during which the accuracy of the scattering positions increases at the expense of the wave speeds accuracy (Fig. 18, red line, path 2). Once the path followed by AST becomes closed to that followed by PAST (namely, when the migration velocities become close to the demigration velocities), AST starts smoothly converging monotonically towards a local minimizer during which accuracy of both velocities and scattering positions improve (Fig. 18, red line, path 3). In contrast, the accuracy of both scatterer positions and wave speeds are monotonically improving during PAST since their updates are tied together by the focusing equations (Fig. 18, black line). In Fig. 18, the first path followed by PAST corresponds to the first multiscale step, where the inversion performs a significant update of the long wavelengths of the velocity model starting from an initial homogeneous model (Fig. 8a,b). During this first phase, both wave speeds and scattering positions are significantly updated. The second path corresponds to the subsequent multiscale steps during which more subtle velocity updates are

performed with however a significant impact on the scattering positions (Fig. 8, c–f). We conclude by noting that PAST not only benefits from a faster convergence compared to AST but also converges towards a more accurate minimizer.

During the Marmousi and real data case studies, we also noted that the most significant improvements in the velocity model obtained with PAST relative to that inferred from AST are shown in the deep part of the subsurface. This failure of AST results from the combined effect of the initialization/localization step and the more limited aperture illumination of the deep subsurface. When the velocities in the starting model tend to be underestimated, the localization step of AST (which fits in a least-squares sense the two slopes and two-way traveltimes) tends to distribute the initial scatterer positions in the subsurface more evenly than PAST (which exactly fits only one slope and two-way traveltimes for scatterer positioning). The more uniform distribution of scatterers in depth performed by the localization step of AST fosters the updating of the full subsurface model in one go along the sensitivity kernels connecting the deep scatterers to the source and receivers (Fig. 4). This all-inclusive reconstruction when combined with a limited aperture illumination of the deep subsurface makes AST more prone to the convergence towards inaccurate deep wave speeds because of the velocity–position trade-off when the initial velocity model is not accurate enough. In contrast, when we use a slow initial velocity model in PAST, the initial scattering positions tend to be concentrated in the upper part of the targeted domain (Fig. 8a). This favours the reconstruction of the shallow velocities during the early iterations of PAST, the scattering positions being progressively moved to their true positions in depth consistently with the velocity updates via the resolution of the focusing equations. In this sense, PAST embeds an implicit layer-stripping or depth continuation strategy, which mitigates the ill-posedness of the slope tomography when combined with the reduced parametrization.

6 CONCLUSION

We revisit the parametrization of slope tomography with the objective of tackling the velocity–position coupling encountered in reflection tomographic methods. We present a remedial formulation that addresses the velocity–position coupling by enforcing physically well-founded constraints under the adjoint-state method. We apply our approach on a synthetic benchmark and a real data while comparing to a more conventional reflection tomography strategy where both the scattering positions and the velocity distribution are jointly updated. We show and elaborate on the superior results in terms of convergence speed and velocity reliability delivered by the presented strategy.

In future works, we will investigate other acquisition-driven strategies like the common-offset case (instead of the common-shot strategy employed in this paper) and examine its possible impact on migration artefacts. We aim to constrain even further our problem on the theoretical fact that scatterers belonging to the same image point, should collapse towards the same position while constructing the correct velocity distribution. Introducing early wide-angle arrivals corresponding to diving waves would complement our slope tomography in terms of illumination and make it more adequate as an FWI starting model building tool.

ACKNOWLEDGEMENTS

REFERENCES

Aki, K. & Lee, W., 1976. Determination of three-dimentional velocity anomalies under a seismic array using first p-arrival times from local earthquakes, *J. geophys. Res*., **81**(4), 4381–4399.

Al-Yahya, K., 1989. Velocity analysis by iterative profile migration, *Geophysics,* **54**(6), 718–729.

Alerini, M., Lambaré, G., Baina, R., Podvin, P. & Bégat, S.L., 2007. Two-dimensional PP/PS-stereotomography: *P*- and *S*-waves velocities estimation from OBC data, *Geophys. J. Int.*., **170,** 725–736.

Alkhalifah, T., 1998. Acoustic approximations for processing in transversely isotropic media, *Geophysics,* **63,** 623–631.

Alkhalifah, T. & Fomel, S., 2010. An eikonal based formulation for travel-time perturbation with respect to the source location, *Geophysics,* **75**(6), T175–T183.

Audebert, F., Nichols, D.E., Rekdal, T., Biondi, B., Lumley, D.E. & Urdaneta, H., 1997. Imaging complex geologic structure with single-arrival Kirchhoff prestack depth migration, *Geophysics,* **62**(5), 1533–1543.

Barbosa, B., Costa, J., Gomes, E. & Schleicher, J., 2008. Resolution analysis for stereotomography in media with elliptic and anelliptic anisotropy, *Geophysics,* **73**(4), R49–R58.

Bauer, A., Schwarz, B. & Gajewski, D., 2017. Utilizing diffractions in wavefront tomography, *Geophysics,* **82**(2), R65–R73.

Billette, F., 1998. *Estimation de macro-modèles de vitesse en sismique réflexion par stéréotomographie,* Ph.D. thesis, Université Paris VII.

Billette, F., Podvin, P. & Lambaré, G., 1998. Stereotomography with automatic picking: application to the Marmousi dataset, in *Expanded Abstracts,* vol. **II,** pp. 1317–1320, Soc. Explor. Geophys.

Billette, F., Le Bégat, S., Podvin, P. & Lambaré, G., 2003. Practical aspects and applications of 2D stereotomography, *Geophysics,* **68,** 1008–1021.

Bishop, T.N., Bube, K.P., Cutler, R.T., Langan, R.T., Love, P.L., Resnick, J.R., Shuey, R.T. & Spinder, D.A., 1985. Tomographic determination of velocity and depth in laterally varying media, *Geophysics,* **50,** 903–923.

Bourgeois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P. & Versteeg, R., 1991. Marmousi, model and data, in *The Marmousi Experience,* pp. 5–16, Eur. Assoc. Explor. Geophys.

Brossier, R., Operto, S. & Virieux, J., 2015. Velocity model building from seismic reflection data by full waveform inversion, *Geophys. Prospect*., **63,** 354–367.

Byrd, R.H., Lu, P. & Nocedal, J., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Stat. Comput.,* **16,** 1190–1208.

Červený, V., 2001. *Seismic Ray Theory,* Cambridge University Press.

Chalard, E., Podvin, P., Lambaré, G. & Audebert, F., 2000. Principles of 3-D stereotomography, in *Extended Abstracts, 62nd Annual Meeting (29 May–2 June 2000, Glasgow),* P–144, Eur. Assoc. Geosci. Eng.

Chauris, H. & Cocher, E., 2017. From migration to inversion velocity analysis, *Geophysics,* **82**(3), S207–S223.

Chauris, H. & Noble, M., 2001. 2D velocity macro model estimation from seismic reflection data by local Differential Semblance Optimization: applications on synthetic and real data, *Geophys. J. Int.,* **144**, 14–26.

Chauris, H., Noble, M., Lambaré, G. & Podvin, P., 2002a. Migration velocity analysis from locally coherent events in 2-D laterally heterogeneous media, part I: theoretical aspects, *Geophysics,* **67**(4), 1202–1212.

Chauris, H., Noble, M., Lambaré, G. & Podvin, P., 2002b. Migration velocity analysis from locally coherent events in 2-D laterally heterogeneous media, part II: applications on synthetic and real data, *Geophysics,* **67**(4), 1213–1224.

Claerbout, J., 1985. *Imaging the Earth's interior,* Blackwell Scientific Publication.

Clément, F., Chavent, G. & Gómez, S., 2001. Migration-based travel-time waveform inversion of 2-D simple structures: a synthetic example, *Geophysics,* **66**, 845–860.

Dummong, S., Meier, K., Gajewski, D. & Hubscher, C., 2008. Comparison of prestack stereotomography and NIP wave tomography for velocity model building: instances from the Messinian evaporites, *Geophysics,* **73**(5), VE291–VE302.

Duveneck, E., 2004. Velocity model estimation with data-derived wavefront attributes, *Geophysics,* **69**(1), 265–274.

Etgen, J., Gray, S.H. & Zhang, Y., 2009. An overview of depth imaging in exploration geophysics, *Geophysics,* **74**(6), WCA5–WCA17.

Farra, V. & Le Bégat, S., 1995. Sensitivity of q*P*-wave traveltimes and polarization vectors to heterogeneity, anisotropy and interfaces, *Geophys. J. Int.,* **121**, 371–384.

Farra, V. & Madariaga, R., 1987. Seismic waveform modeling in heterogeneous media by ray perturbation theory, *J. geophys. Res.,* **92**, 2697–2712.

Farra, V. & Madariaga, R., 1988. Non-linear reflection tomography, *Geophys. J.,* **95**, 135–147.

Fomel, S., Luo, S. & Zhao, H.-K., 2009. Fast sweeping method for the factored eikonal equation, *J. Comput. Phys.,* **228**, 6440–6455.

Gelchinsky, B., Berkovitch, A. & Keydar, S., 1999. Multifocusing homeomorphic imaging Part 1. Basic concepts and formulas, *J. appl. geophys.,* **42**, 229–242.

Golub, G. & Pereyra, V., 2003. Separable nonlinear least squares: the variable projection method and its applications, *Inverse Probl.,* **19**(2), R1, doi:10.1088/0266-5611/19/2/201.

Gosselet, A., Le Bégat, S. & Petersen, S., 2005. Joint slope tomography of borehole transmitted and surface seismic data, in *Expanded Abstracts,* pp. 2577–2580, Soc. Explor. Geophys.

Guillaume, P., *et al.*, 2008. Kinematic invariants: an efficient and flexible approach for velocity model building, in *Advanced Velocity Model Building Techniques for Depth Imaging,* Soc. Explor. Geophys., doi:10.1190/1.3064100

Haber, E., Ascher, U.M. & Oldenburg, D., 2000. On optimization techniques for solving nonlinear inverse problems, *Inverse Probl.,* **16**(5), 1263, doi:10.1088/0266-5611/16/5/309

Hicks, G.J., 2002. Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions, *Geophysics,* **67**, 156–166.

Hocking, R.M., 1988. Regional geology of the northern Carnarvon basin, in *Proc. Petrol. Explor. Soc. Aust. Symp., The North West Shelf, Australia,* pp. 97–114, Petrol. Explor. Soc. Aust., Western Australia

Hu, G., Menke, W. & Powell, C., 1994. Polarization tomography for *P* wave velocity structure in southern California, *J. geophys. Res.,* **99**, 15 245–15 256.

Jin, C. & Zhang, J., 2018. Stereotomography of seismic data acquired on undulant topography, *Geophysics,* **83**(4), U35–U41.

Lambaré, G., 2008. Stereotomography, *Geophysics,* **73**(5), VE25–VE34.

Lavaud, B., Baina, R. & Landa, E., 2004. Post-stack stereotomography: a robust strategy, in *Expanded Abstracts,* p. C22, EAGE, the Netherlands.

Li, Y. & Chauris, H., 2018. Coupling direct inversion to common-shot image-domain velocity analysis, *Geophysics,* **83**(5), R497–R514.

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(5), 645–653.

Métivier, L. & Brossier, R., 2016. The seiscope optimization toolbox: a large-scale nonlinear optimization library based on reverse communication, *Geophysics,* **81**(2), F11–F25.

Montel, J., Guillaume, P., Lambaré, G. & Leblanc, O., 2010. Non-linear slope tomography-extension to maz and waz, in *72nd EAGE Conference and Exhibition incorporating SPE EUROPEC 2010,* EAGE, the Netherlands.

Nag, S., Alerini, M., Duveneck, E. & Ursin, B., 2006. 2-D stereotomography for anisotropic media, in *Expanded Abstracts,* pp. 3305–3308, Soc. Explor. Geophys., Tulsa.

Nguyen, S., Baina, R., Alerini, M. & Lambaré, G., 2008. Stereotomography assisted by migration of attributes, *Geophys. Prospect.*, **56**, 613–625.

Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization,* 2nd edn, Springer.

Operto, S., Xu, S. & Lambaré, G., 2000. Can we image quantitatively complex models with rays? *Geophysics,* **65**(4), 1223–1238.

Paige, C.C. & Saunders, M.A., 1982. ALGORITHM 583 LSQR : sparse linear equations and least squares problems, *ACM Trans. Math. Softw.,* **8**(2), 195–209.

Pavlis, G. & Booker, J., 1980. The mixed discrete-continuous inverse problem: application to the simultaneous determination of earthquake hypocenter and velocity structure, *J. geophys. Res.*, **85**, 4801–4810.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.

Plessix, R.E., Chavent, G. & Roeck, Y.-H.D., 1999. Waveform inversion of reflection seismic data for kinematic parameters by local inversion, *SIAM J. Sci. Comput.,* **20**, 1033–1052.

Prieux, V., Lambaré, G., Operto, S. & Virieux, J., 2013. Building starting model for full waveform inversion from wide-aperture data by stereotomography, *Geophys. Prospect.*, **61**, 109–137.

Qian, J. & Symes, W., 2002. An adaptive finite-difference method for traveltimes and amplitudes, *Geophysics,* **67**, 167–176.

Riabinkin, L.A., 1957. *Fundamentals of Resolving Power of Controlled Directional Reception (CDR) of Seismic Waves,* Vol. **14**, pp. 36–60, Soc. Explor. Geophys., Tulsa (translated and paraphrased from Prikladnaya, 16, 3–36).

Rieber, F., 1936. A new reflection system with controlled directional sensitivity, *Geophysics,* **1**, 97–106.

Sava, P. & Biondi, B., 2004. Wave-equation migration velocity analysis. i. theory, *Geophys. Prospect.*, **52**(6), 593–606.

Shen, P., Symes, W. & Stolk, C., 2003. Differential semblance velocity analysis by wave-equation migration, in *SEG Technical Program Expanded Abstracts 2003,* pp. 2132–2135, Soc. Explor. Geophys., Tulsa.

Soubaras, R. & Whiting, P., 2011. Variable depth streamer — the new broadband acquisition system in *SEG Technical Program Expanded Abstracts, Vol. 30,* pp. 4349–4353, Soc. Explor. Geophys., Tulsa.

Spencer, C. & Gubbins, D., 1980. Travel time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media, *Geophys. J. R. Astron. Soc.,* **63**, 95–116.

Stork, C. & Clayton, R.W., 1985. Iterative tomographic and migration reconstruction of seismic images, in *SEG Technical Program Expanded Abstracts 1985,* pp. 610–613, Soc. Explor. Geophys., Tulsa.

Stork, C. & Clayton, R.W., 1986. Analysis of the resolution between ambiguous velocity and reflector position for traveltime tomography, in *SEG Technical Program Expanded Abstracts 1986,* pp. 545–550, Soc. Explor. Geophys., Tulsa.

Sword, C.H., 1987. *Tomographic Determination of Interval Velocities from Reflection Seismic Data: The Method of Controlled Directional Reception,* Ph.D. thesis, Stanford University.

Symes, W.W. & Carazzone, J.J., 1991. Velocity inversion by differential semblance optimization, *Geophysics,* **56**, 654–663.

Taillandier, C., Noble, M., Chauris, H. & Calandra, H., 2009. First-arrival travel time tomography based on the adjoint state method, *Geophysics,* **74**(6), WCB1–WCB10.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation,* Elsevier.

Tavakoli, F.B., Operto, S., Ribodetti, A., Sambolian, S. & Virieux, J., 2018. Anisotropic first arrival slope and traveltime tomography (FASTT), in *Expanded Abstracts, 80th Annual EAGE Meeting,* EAGE, Copenhagen, doi: 10.3997/2214-4609.201801387

Tavakoli, F.B., Operto, S., Ribodetti, A. & Virieux, J., 2019. Matrix-free anisotropic slope tomography: theory and application, *Geophysics,* **84(1)**, R35–R57.

Tavakoli, F.B., Ribodetti, A., Virieux, J. & Operto, S., 2015. An iterative factored eikonal solver for TTI media, in *SEG Technical Program Expanded Abstracts 2015,* Vol. **687**, pp. 3576–3581, Soc. Explor. Geophys., Tulsa.

Tavakoli, F.B., Operto, S., Ribodetti, A. & Virieux, J., 2017. Slope tomography based on eikonal solvers and the adjoint-state method, *Geophys. J. Int.,* **209**(3), 1629–1647.

ten Kroode, A.P.E., Smit, D.J. & Verdel, A.R., 1994. Linearized inverse scattering in the presence of caustics, in *Expanded Abstracts,* SPIE, Bellingham.

Valensi, R., Baina, R. & Duprat, V., 2017. Reflection waveform inversion method: solutions to the reflectivity-background coupling problem and consequences on the convergence, in *79th EAGE Conference and Exhibition 2017 - Workshops,* EAGE Publications BV, Copenhagen.

Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.

Waheed, U.B., Yarman, C.E. & Flagg, G., 2014. An iterative fast sweeping based eikonal solver for tilted orthorhombic media, in *Expanded Abstracts,* pp. 480–485, Soc. Explor. Geophys., Tulsa.

Wu, Z. & Alkhalifah, T., 2015. Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion, *Geophysics,* **80**(6), R317–R329.

Xu, S., Chauris, H., Lambaré, G. & Noble, M., 2001. Common-angle migration: a strategy for imaging complex media, *Geophysics,* **66,** 1877–1894.

Xu, S., Wang, D., Chen, F., Lambaré, G. & Zhang, Y., 2012. Inversion on reflected seismic wave, in *SEG Technical Program Expanded Abstracts 2012,* pp. 1–7, Soc. Explor. Geophys., Tulsa.

Yang, K., Shao, W.-D., Xing, F.-Y. & Xiong, K., 2018. Stereotomography in triangulated models, *Geophys. J. Int.,* **214**(2), 1018–1040.

Zhao, H., 2005. A fast sweeping method for eikonal equations, *Math. Comput.,* **74,** 603–627.

Zhou, W., Brossier, R., Operto, S. & Virieux, J., 2015. Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation, *Geophys. J. Int.,* **202**(3), 1535–1554.

Zhou, W., Brossier, R., Operto, S., Virieux, J. & Yang, P., 2018. Velocity model building by waveform inversion of early arrivals and reflections: a 2D ocean-bottom-cable study with gas cloud effects, *Geophysics,* **83**(2), R141–R157.

APPENDIX A: ALGORITHMIC ASPECT OF SOLVING THE FOCUSING EQUATIONS

In the following, we describe the manner of solving the focusing eq. (12). We first aim to define an isochrone by setting a threshold on traveltime difference between the observed traveltime associated with a scatterer and the computed counterpart at each grid node in the two-way traveltime map of the corresponding source–receiver couple. The threshold is proportional to the grid spacing and the average velocity in the model. After generating the logical map containing the candidate positions belonging to the isochrone, we proceed to the second step related to the slope. We slide a 3×3 window centred on the positions spanned by the isochrone, we then examine the sign of the difference between the slope value at the four corners of the sliding window and the observed slope of the concerned scattering point (same finite difference scheme as the one describe by eq. 22). Two scenarios are possible, the first where the sign of the residuals at the four corners is the same while the second would reveal the contrary. If the first scenario occurs, the sliding window moves to the next candidate position without any further evaluation. On the other hand, in occurrence of the second, further evaluation of the slopes is done on the remaining five nodes of the sliding window in order to determine the quadrant that is pointing towards the optimal position. This process flags some parts of the isochrone as being inadequate and accordingly the sliding window moves to the most plausible part of the isochrone without exhaustively examining all of its parts. Once the sliding window locks on the most susceptible position, we interpolate very finely in and around it using the same sampling operator of eq. (22) and then examine both the traveltime and slope residual. We proceed with a thorough grid search and finally retain the position that validates perfectly the focusing equations or the optimal one that fits both slope and traveltime to a certain order of accuracy. The chosen grid search strategy for the kinematic migration is efficient and well optimized, the cost is elementary, scales linearly with the number of processors and overall its cost is compensated throughout the totality of the inversion process. Other optimization alternatives are possible, most would involve scaling on the data or are simply less efficient.

APPENDIX B: EIKONAL-BASED SLOPE TOMOGRAPHY GRADIENT WITH FRéCHET DERIVATIVES

In this appendix, we develop the expression of the Fréchet derivative matrix associated with PAST by adapting the ray-based formalism of Chauris *et al.* (2002a, their appendix B) to our method.

We seek to compute the partial derivative of $p_{s,n_{s,r}}$ with respect to a subsurface parameter m_l. Differentiation of data-domain attributes $(p_{s,n_{s,r}}, T_{s,r,n_{s,r}}, p_{r,n_{s,r}})$ parametrizing a locally coherent event with respect to the corresponding scatterer coordinates $(x_{n_{s,r}}, z_{n_{s,r}})$ and the subsurface parameter m_l gives the following system

$$
\begin{pmatrix} \mathrm{d}p_{s,n_{s,r}} \\ \mathrm{d}T_{s,r,n_{s,r}} \\ \mathrm{d}p_{r,n_{s,r}} \end{pmatrix} = \left(\frac{\partial \left(p_{s,n_{s,r}}, T_{s,r,n_{s,r}}, p_{r,n_{s,r}} \right)}{\partial \left(m_l, x_{n_{s,r}}, z_{n_{s,r}} \right)} \right) \cdot \begin{pmatrix} \delta m_l \\ \delta x_{n_{s,r}} \\ \delta z_{n_{s,r}} \end{pmatrix}.
\tag{B1}
$$

Since $T_{s,r,n_{s,r}}$ and $p_{r,n_{s,r}}$ are forced to be equal to the measurements $T^{*}_{s,r,n_{s,r}}$ and $p^{*}_{r,n_{s,r}}$ in PAST, $dT_{s,r,n_{s,r}} = 0$ and $dp_{r,n_{s,r}} = 0$. Therefore, the system becomes

$$\begin{pmatrix} dp_{s,n_{s,r}} \\ 0 \\ 0 \end{pmatrix} = \left(\frac{\partial \left(p_{s,n_{s,r}}, T_{s,r,n_{s,r}}, p_{r,n_{s,r}} \right)}{\partial \left(m_l, x_{n_{s,r}}, z_{n_{s,r}} \right)} \right) \cdot \begin{pmatrix} \delta m_l \\ \delta x_{n_{s,r}} \\ \delta z_{n_{s,r}} \end{pmatrix}. \tag{B2}$$

Solving this system for δm_l with Cramer's rule gives the partial derivative of $p_{s,n_{s,r}}$ with respect to m_l

$$\frac{\partial p_{s,n_{s,r}}}{\partial m_l} = \frac{\det \left| \dfrac{\partial (p_{s,n_{s,r}}, T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial (m_l, x_{n_{s,r}}, z_{n_{s,r}})} \right|}{\det \left| \dfrac{\partial (T_{s,r,n_{s,r}}, p_{r,n_{s,r}})}{\partial (x_{n_{s,r}}, z_{n_{s,r}})} \right|}. \tag{B3}$$

Repeating this for all the parameters m_l of the subsurface domain builds the full sensitivity matrix. The Fréchet derivative is similar to that of Chauris *et al.* (2002a, eq. 39) except that our eikonal-based forward problem leads to more compact data and model spaces. The denominator in eq. (B3) is indeed the same as the one involved in the expression of the adjoint-state variables $\xi_{r,n_{s,r}}$ and $\mu_{s,r,n_{s,r}}$, eqs (18) and (19), and defines the reflection imaging condition. The numerator would require the computation of the partial derivative of $p_{s,n_{s,r}}$ with respect to m_l, which would reveal prohibitively expensive, while the partial derivatives of $p_{s,n_{s,r}}$, $T_{s,r,n_{s,r}}$ and $p_{r,n_{s,r}}$ with respect to $x_{n_{s,r}}$ and $z_{n_{s,r}}$ can be easily inferred from the source and receiver traveltime maps (see eq. 17). Note also how the determinant of the 3 x 3 Jacobian in the numerator of eq. (B3) is broken down into the determinants of two 2 x 2 Jacobians in the adjoint-state approach, eqs (18) and (19), since we avoid the explicit computation of the derivative of $p_{s,n_{s,r}}$ with respect to m_l.