WAVEFRONT TOMOGRAPHY BY DYNAMIC FOCUSING

P. Znak, B. Kashtan, and D. Gajewski

email: *pavel.znak@uni-hamburg.de*

keywords: *tomography, curvatures, geometrical spreading, focusing*

ABSTRACT

A reliable and fast automatic workflow for smooth velocity model building is of interest for subsequent depth migration and as a source of initial models for full-waveform inversion. At the same time, seismic data enhancement by means of stacking is widely performed. Common-reflection-surface (CRS) stack results in kinematic wavefield attributes, which locally characterize the wavefront. We propose a novel wavefront curvature based post-stack tomography method valid for reflection, diffraction and passive seismic data inversion. The basis of the new approach is an overall minimizing of diffracted waves geometrical spreading at source positions and emergence time. We call this procedure dynamic focusing. During the optimization the wavefield attributes remain constant and serve as initial conditions and travel time for kinematic and dynamic ray tracing. The natural parametrization of the inverse problem by velocity model as the only unknown significantly decreases tomographic matrix dimension and improves the data-unknowns ratio. After a velocity model is retrieved reflectors and diffractors are localized by ray tracing. Since deriving and coding the conventional wavefront tomography is cumbersome even in the isotropic case, the reduction of the Fréchet derivatives number, which takes place in our approach, is an attractive feature if generalization to anisotropy is desired. Both Fréchet derivatives and adjoint-state method formulae for computing the gradient were derived. The algorithm combined with the L-BFGS-B quasi-Newton solver was tested on a full of diffracted energy marine dataset.

INTRODUCTION

Seismic reflection and diffraction multiparameter data stacking is a conventional and powerful processing tool. It universally serves for the signal-to-noise ratio enhancement both for the stack section (Mayne, 1962; Mann et al., 1999; Jäger et al., 2001) and pre-stack data (Baykulov and Gajewski, 2009), crucially saves the depth migration computing time and produces physically meaningful wavefield attributes (Hubral, 1983). The quality of a stack section depends on the choice of stacking travel time based operator. Historically the first, normal move-out stack for common mid-point (CMP) gathers (Mayne, 1962) was generalized by common-reflection-surface (CRS) operator (Mann et al., 1999; Jäger et al., 2001) designed to utilize the redundancy of the seismic records. Multifocusing (Gelchinsky et al., 1999) is known to be efficient in smooth environments with curved interfaces, whereas CRS is robust to overburden heterogeneities. Recently, an implicit CRS (iCRS) method accounting for the reflector curvature and combining the advantages of CRS and multifocusing was proposed (Schwarz et al., 2014).

Kinematic wavefield attributes represent an extra output of the seismic data stacking during the coherence analysis. For the CRS operator they describe two hypothetical waves: a wave emerging from a zero-offset ray normal incidence point and an exploding reflector wave (Hubral, 1983). Namely, four attributes are zero-offset ray travel time, horizontal component of slowness, curvature of the normal incidence point wave wavefront and curvature of the exploding reflector wave wavefront. For the point diffractor the kinematic wavefield attributes describe the real wavefront and these curvatures coincide. This property yields a criterion for diffraction separation (Dell and Gajewski, 2011).

The idea of utilizing kinematic wavefield attributes representing an extra stacking output as input data for a fast smooth velocity model inversion was first proposed in a paper of Duveneck (2004) and improved by Dell et al. (2014). Recently, this method of zero-offset wavefront tomography has found its application to passive seismic source localization (Schwarz et al., 2016) and diffraction imaging (Bauer et al., 2017).

The inverse problem was formulated similarly to stereotomography (Lambaré, 2008) with velocity model, diffractors and scattering angles as unknowns and objective function minimizing the attributes and mid-point coordinates misfits. The number of unknowns depends on the number of data points. When a new pick is added to the dataset four numbers are added to the data vector but automatically three new unknowns arise. Seismic tomography is ill-conditioned and requires sufficient regularization which smooths the velocity model and reduces resolution (Hansen, 1998; Costa et al., 2008). Usually, the smaller is the size of tomographic matrix the smaller regularization is needed. Moreover, a stable and unique inversion implies the amount of data significantly exceeding the amount of unknowns, which is not the case in this way formulated wavefront tomography. In contrast to the pre-stack stereotomography the wavefront tomography data contains travel times for all involved rays. To fully exploit this fact we developed a novel kinematic wavefield attributes inversion workflow and tested it with a full of diffracted energy marine dataset from the Levantine Basin (Netzeband et al., 2006).

THEORY AND METHOD

The hyperbolic zero-offset CRS stacking operator

$$t^2(\triangle x_m, h) = (t_0 + 2p_0 \triangle x_m)^2 + 2t_0(\hat{M}^N \triangle x_m^2 + \hat{M}^{NIP} h^2) \tag{1}$$

results in zero-offset travel time t_0, horizontal component of slowness p_0, second derivative of normal incidence point (NIP) wave travel time $\hat{M}^{NIP}$ and second derivative of normal (N) exploding reflector wave travel time $\hat{M}^N$. If the surface velocity is locally constant $\hat{M} = \cos^2 \alpha_0 M$, where α_0 is an angle at which the zero-offset ray arrives to the surface and M is a ratio of projection of slowness derivative along the wavefront to the wavefront P and geometrical spreading Q (Cervený, 2001). It is related to the wavefront curvature $\kappa = vM$. If a point diffractor is considered the kinematic wavefield attributes describe the actual wavefront and $\hat{M}^N = \hat{M}^{NIP}$.

An objective function introduced by Duveneck (2004)

$$J(\mathbf{m}) = \sum_{i=1}^{N_{data}} (t_i - t_i(\mathbf{m}_i))^2 + \sum_{i=1}^{N_{data}} (x_i - x_i(\mathbf{m}_i))^2 + \sum_{i=1}^{N_{data}} (p_i - p_i(\mathbf{m}_i))^2 + \sum_{i=1}^{N_{data}} (\hat{M}_i - \hat{M}_i(\mathbf{m}_i))^2 \tag{2}$$

is a sum of squared attributes misfits over all the automatically picked N_{data} data points. A vector of unknowns $\mathbf{m}$ comprises diffractors coordinates x_i^d, z_i^d and zero-offset rays emergence angles α_i^d for each pick $i = 1, 2, \ldots, N_{data}$ together with coefficients v_{ij} defining a smooth velocity model in terms of B-splines

$$v = \sum_{j=1}^{N_x} \sum_{k=1}^{N_z} v_{jk} \beta_j(x) \beta_k(z) \tag{3}$$

with $N_x \times N_z$ nodes, $\mathbf{m}_i = (x_i^d, z_i^d, \alpha_i^d, v)$. The defined velocity model has a slight spatial non-equality. Due to a lack of the B-splines outside of the model velocity decays to the model boundaries in the case of equal v_{jk}. We achieve a constant velocity model with equal v_{jk} coefficients by nodes sequence expansion beyond the model boundaries. The external nodes values are chosen to coincide with the boundary coefficients. The inverse problem has $4N_{data}$ data points and $3N_{data} + N_x N_z$ unknowns. The number of unknowns depends on the data points number. At best the N_{data} picks number significantly exceeds the number of basis functions nodes. This leads to a ratio of the number of data points and the number of unknowns equal to $\frac{4}{3}$. Other cases with higher velocity sampling leads to even smaller ratio close to one. We aim to have a free choice of unknowns number for a given data amount.

The first idea that comes to mind is to rearrange the available data replacing the previous objective function (2) with one minimizing only wavefronts curvatures. If x, p and t are fixed the objective function

$$J(v) = \sum_{i=1}^{N_{data}} (\hat{M}_i - \hat{M}_i(\mathbf{m}_i(v)))^2 \tag{4}$$

becomes a function of only velocity model. The coordinates of diffractors and the zero-offset ray emergence angles are determined by velocity model with time reversal ray tracing as soon as the values x, p and t are given, $\mathbf{m}_i(v) = (x_i^d(v), z_i^d(v), \alpha_i^d(v), v)$. $\hat{M}$ turns out to be a composition of reverse time kinematic ray tracing to a diffractor and forward dynamic ray tracing from the diffractor to the surface. Hence, to get a gradient of this objective function one needs a chain rule for Fréchet derivatives. All the partial derivatives could be interpreted within kinematic and dynamic ray tracing perturbation theory (Farra and Madariaga, 1987). However, we find this approach cumbersome and present another natural solution which preserves all the advantages and doesn't require a hard derivation. The idea is to minimize a third-party quantity at the reflector (diffractor) position while keeping all the kinematic wavefield attributes fixed as initial conditions and travel time for kinematic and dynamic ray tracing.

Dynamic focusing

We refer to dynamic focusing as a procedure of simultaneous minimizing of geometrical spreading of all the diffracted and normal incidence point waves presented in the kinematic wavefield attributes dataset. Wherein the kinematic attributes t, M, p are considered as fixed at x experimental quantities. Since diffractors and normal incidence points are kinematically equivalent to the point sources, for the true velocity model geometrical spreading vanishes when is back propagated up to arrival time. This imaging principle is valid for the passive seismic sources as well.

To be solved in depth the dynamic ray tracing system (Cervený, 2001)

$$\frac{d}{dt}\begin{pmatrix} Q \\ P \end{pmatrix} = S\begin{pmatrix} Q \\ P \end{pmatrix}, \quad S = \begin{pmatrix} 0 & v^2 \\ -\frac{1}{v}\frac{\partial^2 v}{\partial q^2} & 0 \end{pmatrix}. \tag{5}$$

requires the initial conditions for Q and P. Unfortunately, they are not presented in the set of CRS attributes. The only available after processing dynamic quantity is the related to wavefront curvature value M. Although dynamic tracing of M is theoretically possible by solving the Riccati equation, it fails due to unlimited growth in the vicinity of point sources and caustics. It is known that geometrical spreading Q is defined on a ray only up to a certain constant quotient determined by rays fan parametrization. If the geometrical spreading is defined somehow, e.g., by emergence angle parametrization, with the help of M value at a ray point we are able to retrieve the geometrical spreading on the whole ray up to a constant quotient. This is done by choosing the initial conditions satisfying

$$\frac{P(0)}{Q(0)} = M. \tag{6}$$

Since there exist an infinite number of such choices, we fix one with a condition for geometrical spreading to be normalized at the surface. All the attributes picks are independent. Hence, in order to simplify the notations we assume the corresponding waves emerging at the same time $t = 0$ and arriving to the surface at coordinate x and time $\tau(x)$, which is known from the experiment. A propagator of dynamic ray tracing system (5)

$$\Pi(x; t, \tau(x)) = \begin{pmatrix} Q^{(1)}(x; t, \tau(x)) & Q^{(2)}(x; t, \tau(x)) \\ P^{(1)}(x; t, \tau(x)) & P^{(2)}(x; t, \tau(x)) \end{pmatrix}, \quad \Pi(x; \tau(x), \tau(x)) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{7}$$

has to be traced back in time ($t < \tau(x)$) to the subsurface. We define normalized dynamic ray tracing quantities as a special linear combination of its columns

$$\begin{pmatrix} \hat{Q} \\ \hat{P} \end{pmatrix}(x, t) = \Pi(x; t, \tau(x)) \begin{pmatrix} 1 \\ M(x) \end{pmatrix}, \tag{8}$$

where $M(x)$ is the experimental wavefront curvature related attribute at x coordinate. $Q^{(1)}$ and $P^{(1)}$ are multiplied by one with the units of M. Therefore, the units of $\hat{Q}$ coincide with the units of P. As a linear combination of solutions the normalized quantities vector satisfies the dynamic ray tracing system (5). Indeed, for normalized dynamic quantities we obtain $\hat{Q}(x, \tau(x)) = 1$ and $\hat{P}(x, \tau(x)) = M(x)$. Therefore,

$$\frac{\hat{P}(x, \tau(x))}{\hat{Q}(x, \tau(x))} = \frac{P(x, \tau(x))}{Q(x, \tau(x))} = M(x), \tag{9}$$

where $Q(x,t)$ and $P(x,t)$ are the actual dynamic quantities along the ray. Due to linearity of the system the actual Q,P pair and the pair of normalized $\hat{Q},\hat{P}$ are proportional at any travel time with the quotient equal to the actual geometrical spreading value at the surface

$$Q(x,t) = Q(x,0)\hat{Q}(x,t), \quad P(x,t) = Q(x,0)\hat{P}(x,t). \tag{10}$$

Geometrical spreading for a given travel time and M value can be retrieved only up to a constant. However, if at some travel time the actual geometrical spreading Q vanishes the normalized geometrical spreading $\hat{Q}$ necessarily does the same. In particular at the source positions

$$\hat{Q}(x,0) = Q(x,0) = 0. \tag{11}$$

In accordance to this we introduce an objective functional for the dynamic focusing as a sum of squared normalized geometrical spreading propagated back up to one-way travel time over all the data picks

$$J = \frac{1}{2}\sum_{i=1}^{N_{data}} \hat{Q}^2(x_i,0). \tag{12}$$

Finally, the wavefront tomography is formulated as inversion in velocity model as the only unknown with N_{data} data dimension and $N_x \times N_z$ - dimensional unknowns space. The tomographic matrix block correspondent to a pick at x_i of the dynamic focusing

$$\frac{\partial \hat{Q}_i}{\partial v_1} \quad \frac{\partial \hat{Q}_i}{\partial v_2} \quad \cdots \quad \frac{\partial \hat{Q}_i}{\partial v_{NxNz}} \tag{13}$$

and of the broad functional approach

$$\begin{array}{cccccc}
\frac{\partial t_i}{\partial x_i^d} & \frac{\partial t_i}{\partial z_i^d} & \frac{\partial t_i}{\partial \alpha_i^d} & \frac{\partial t_i}{\partial v_1} & \frac{\partial t_i}{\partial v_2} & \cdots \quad \frac{\partial t_i}{\partial v_{N_x N_z}} \\
\frac{\partial x_i}{\partial x_i^d} & \frac{\partial x_i}{\partial z_i^d} & \frac{\partial x_i}{\partial \alpha_i^d} & \frac{\partial x_i}{\partial v_1} & \frac{\partial x_i}{\partial v_2} & \cdots \quad \frac{\partial x_i}{\partial v_{N_x N_z}} \\
\frac{\partial p_i}{\partial x_i^d} & \frac{\partial p_i}{\partial z_i^d} & \frac{\partial p_i}{\partial \alpha_i^d} & \frac{\partial p_i}{\partial v_1} & \frac{\partial p_i}{\partial v_2} & \cdots \quad \frac{\partial p_i}{\partial v_{N_x N_z}} \\
\frac{\partial M_i}{\partial x_i^d} & \frac{\partial M_i}{\partial z_i^d} & \frac{\partial M_i}{\partial \alpha_i^d} & \frac{\partial M_i}{\partial v_1} & \frac{\partial M_i}{\partial v_2} & \cdots \quad \frac{\partial M_i}{\partial v_{N_x N_z}}
\end{array} \tag{14}$$

illustrate the redundancy of the latter one.

Fréchet derivatives and adjoint-state method

Fréchet derivative of the functional (gradient) can be expressed through the Fréchet derivatives of the state variables

$$\triangle J = \sum_{i=1}^{N_{data}} \hat{Q}(x_i,0)\triangle\hat{Q}(x_i,0). \tag{15}$$

From the perturbation theory of dynamic ray tracing (Farra and Madariaga, 1987) we know that perturbations of dynamic quantities satisfies the inhomogeneous system

$$\frac{d}{dt}\begin{pmatrix} \triangle\hat{Q} \\ \triangle\hat{P} \end{pmatrix} = S\begin{pmatrix} \triangle\hat{Q} \\ \triangle\hat{P} \end{pmatrix} + \triangle S\begin{pmatrix} \hat{Q} \\ \hat{P} \end{pmatrix}, \tag{16}$$

$$\triangle S = S_{\triangle q,\triangle p} + S_{\triangle v}, \tag{17}$$

$$S_{\triangle q,\triangle p} = \begin{pmatrix} 2v\frac{\partial v}{\partial q}\triangle p & 2v\frac{\partial v}{\partial q}\triangle q \\ \left(\frac{3}{v^2}\frac{\partial v}{\partial q}\frac{\partial^2 v}{\partial q^2} - \frac{1}{v}\frac{\partial^3 v}{\partial q^3}\right)\triangle q & -2v\frac{\partial v}{\partial q}\triangle p \end{pmatrix}, \quad S_{\triangle v} = \begin{pmatrix} 0 & v\triangle v \\ -\frac{\partial^2}{\partial q^2}\left(\frac{\triangle v}{v}\right) + \frac{\triangle v}{v^2}\frac{\partial^2 v}{\partial q^2} & 0 \end{pmatrix}, \tag{18}$$

where $\triangle q(t)$ and $\triangle p(t)$ are ray-centered coordinate and slowness of the perturbed ray, which in turn have to be evaluated by

$$\frac{d}{dt}\begin{pmatrix} \triangle q \\ \triangle p \end{pmatrix} = S\begin{pmatrix} \triangle q \\ \triangle p \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{\triangle v}{v^2}\frac{\partial v}{\partial q} - \frac{1}{v}\frac{\partial \triangle v}{\partial q} \end{pmatrix}. \tag{19}$$

The initial conditions $\triangle\hat{Q}(x,\tau(x))\ \triangle\hat{P}(x,\tau(x))$ are zero because $\hat{Q}$ and $\hat{P}$ are defined by the elements of propagator. And the propagator is equal to identity matrix at $t=\tau(x)$ regardless of the velocity model. We fix the arrival point of the ray and, hence, $\triangle q(x,\tau(x))$ is also zero. However, $\triangle p(x,\tau(x))$ differs. p_x is assumed to be known from the experiment and also fixed. Accounting for the first order perturbation of the eikonal equation at the point of arrival

$$(\mathbf{p},\triangle\mathbf{p}) = -\frac{\triangle v}{v^3} \tag{20}$$

together with the transform of perturbed slowness $\triangle\mathbf{p}$ from Cartesian to ray-centered coordinate system

$$\begin{pmatrix} -\frac{\triangle v}{v^2} \\ \triangle p \end{pmatrix} = \begin{pmatrix} vp_x & vp_z \\ -vp_z & vp_x \end{pmatrix} \begin{pmatrix} 0 \\ \triangle p_z \end{pmatrix}, \tag{21}$$

yields

$$\triangle p(x,\tau(x)) = -\frac{\triangle v}{v_0^2}\frac{p_x}{p_z}. \tag{22}$$

The inhomogeneous system (16) is solved with the propagator of homogeneous one (see, e.g, Gilbert and Backus, 1966; Bellman, 1997)

$$\triangle\hat{Q}(x,0) = \begin{pmatrix} Q^{(1)}(x;0,\tau(x)) \\ Q^{(2)}(x;0,\tau(x)) \end{pmatrix}^T \int\limits_{\tau(x)}^{0} \Pi^{-1}(x;t',\tau(x))\triangle S(x,t') \begin{pmatrix} \hat{Q}(x,t') \\ \hat{P}(x,t') \end{pmatrix} dt'. \tag{23}$$

We also derived a formula (see the Appendix) for the gradient in form of the adjoint-state method (Plessix, 2006; Chavent, 2010):

$$\triangle J = \sum_{i=1}^{N_{data}} \int\limits_{t(x)}^{0} \left(-\hat{Q}(x_i,t') - \lambda_P(x_i,t')\,\lambda_Q(x_i,t')\right) \triangle S(x,t') \begin{pmatrix} \hat{Q}(x_i,t') \\ \hat{P}(x_i,t') \end{pmatrix} dt', \tag{24}$$

where adjoint-state variables λ_Q and λ_P are uniquely defined being a solution of inhomogeneous system

$$\frac{\partial}{\partial t}\begin{pmatrix} \lambda_Q \\ \lambda_P \end{pmatrix} = S\begin{pmatrix} \lambda_Q \\ \lambda_P \end{pmatrix} + v^2\begin{pmatrix} \hat{Q} \\ -\hat{P} \end{pmatrix} \tag{25}$$

with zero initial conditions at emergence time $\lambda_Q(x,0)=0$, $\lambda_P(x,0)=0$. The dynamic ray tracing system is self-adjoint.

$$\begin{pmatrix} \lambda_Q \\ \lambda_P \end{pmatrix}(x,t) = \Pi(x;t,0)\int\limits_0^t v^2(x,t')\Pi^{-1}(x;t',0)\begin{pmatrix} \hat{Q} \\ -\hat{P} \end{pmatrix}(x,t')dt'. \tag{26}$$

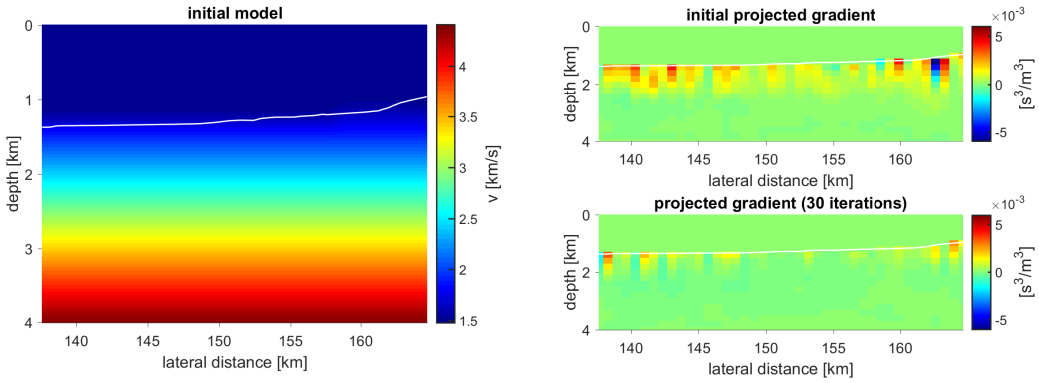
The property

$$\Pi(x;t,0) = \Pi(x;t,\tau(x))\Pi(x;\tau(x),0) \tag{27}$$

provides a way to compute the adjoint-state variables

$$\begin{pmatrix} \lambda_Q \\ \lambda_P \end{pmatrix}(x,t) = \Pi(x;t,\tau(x))\int\limits_0^t v^2(x,t')\Pi^{-1}(x;t',\tau(x))\begin{pmatrix} \hat{Q} \\ -\hat{P} \end{pmatrix}(x,t')dt'. \tag{28}$$

The adjoint-state method formula (24) assumes an integration of the scalar function, whereas the gradient expressed through the Fréchet derivatives (15, 23) needs an integration of the vector. However, the elements of this vector are computed not independently because the $\triangle S$ matrix is prepared for both. And the additional integration during the adjoint-state system solution may cancel the adjoint-state method performance advantage. Indeed, we haven't noticed a significant difference in the performance comparing the methods for 21×41 4th order B-splines nodes grid. But increasing the nodes number in the ray propagation direction and the basis function support (e.g., by taking the B-splines of higher order) could make this advantage significant. Probably, this also can make sense with transition to 3D similar to the stereotomography (Plessix, 2006).



(a) **(b)**

Figure 1: (a) Initial constant gradient overburden. (b) Objective function projected gradients.

FIELD DATA EXAMPLE

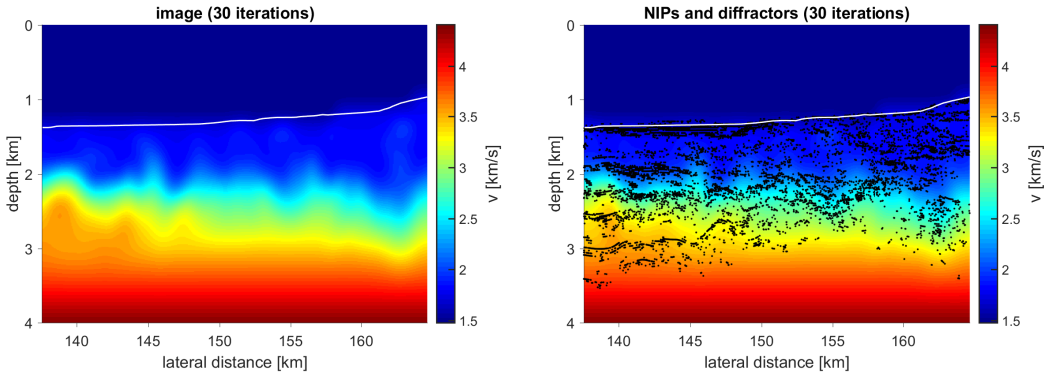
A marine field dataset provided by TGS was already processed by Bauer et al. (2017) with diffraction and reflection simultaneous inversion by means of Duveneck (2004) workflow. The full of diffracted energy data was collected in the Levantine Basin of Mediterranean sea in a strongly heterogeneous part with salt rollers (Netzeband et al., 2006). 5525 diffraction picks in addition to 6430 reflection picks were inverted by dynamic focusing in the frame of 21×41 4th order B-spline nodes grid with 675 m nodes spacing in horizontal direction and 200 m nodes spacing in the vertical direction. To perform convergence we used limited-memory constrained Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) algorithm (Byrd et al., 1995) starting with a model from Figure 1(a). The initial velocity linearly increases from 1.8 km/s for sediments at the seabed (marked with a white line) to 4.4 km/s for salt at the lower boundary of the model. The constraints for the algorithm were the following: the nodes values over the seabed were fixed (lower and upper limits equal to the water velocity 1480 m/s), lower and upper limits for the subsurface nodes values coincide with the limits of the initial constant gradient. We applied no regularization (no additional terms in the functional and gradient like in Tikhonov or total variation approach). A comparison of initial and final projected gradients is an illustration for the convergence process (Figure 1(b)). As a measure of functional variation with slight increment in velocity these plots outline the domains of maximum convergence impact. The result of inversion is presented in Figure 2(a). An interface between the seabed sedimentary layer and salt is in good agreement with the image obtained by Bauer et al. (2017) for the same dataset. The diffractors and normal incidence points were finally traced to their positions and superimposed on the retrieved velocity image (Figure 2(b)).

CONCLUSIONS

A novel workflow for smooth velocity model building by inversion of kinematic wavefield attributes is proposed and tested with marine field dataset. Concerning the ratio and independence of the data points number and the number of unknowns the method is more attractive than the conventional one. Both the Fréchet derivatives and the adjoint-state method gradients were derived. In application to the Levantine Basin dataset the retrieved tomographic image highly correlate with the previous studies. We consider the new workflow to be promising and a comprehensive comparison with the conventional wavefront tomography with the same data, grid spacing and regularization as the next step in our research.

ACKNOWLEDGMENTS

The authors appreciate the support of the *Wave Inversion Technology (WIT) Consortium* sponsors. We thank TGS for providing the field dataset and members of the applied seismic group of Hamburg University: Jan Walda for preparing the kinematic wavefield attributes, Ivan Abakumov and Sergius Dell for the assistance and fruitful discussions.



(a) **(b)**

Figure 2: (a) Retrieved tomographic image. (b) Localized normal incidence points and diffractors.

REFERENCES

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

Baykulov, M. and Gajewski, D. (2009). Prestack seismic data enhancement with partial common-reflection-surface (CRS) stack. *Geophysics*, 74(3):V49–V58.

Bellman, R. (1997). *Introduction to matrix analysis*. Society for Industrial and Applied Mathematics, 2nd edition.

Byrd, R., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Cervený, V. (2001). *Seismic ray theory*. Cambridge university press.

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

Costa, J., Silva, F., Gomes, E., Schleicher, J., Melo, L., and Amazonas, D. (2008). Regularization in slope tomography. *Geophysics*, 73(5):VE39–VE47.

Dell, S. and Gajewski, D. (2011). Common-reflection-surface-based workflow for diffraction imaging. *Geophysics*, 76(5):S187–S195.

Dell, S., Gajewski, D., and Tygel, M. (2014). Image-ray Tomography. *Geophysical Prospecting*, 62(3):413–426.

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

Farra, V. and Madariaga, R. (1987). Seismic waveform modeling in heterogeneous media by ray perturbation theory. *Journal of Geophysical Research*, 92(B3):2697–2712.

Gelchinsky, B., Berkovitch, A., and Keydar, S. (1999). Multifocusing homeomorphic imaging, Part 1: Basic concepts and formulae. *Journal of Applied Geophysics*, 42(3):229–242.

Gilbert, F. and Backus, G. (1966). Propagator matrices in elastic wave and vibration problems. *Geophysics*, 31(2):326–332.

Hansen, P. (1998). *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Society for Industrial and Applied Mathematics.

Hubral, P. (1983). Computing true amplitude reflections in a laterally inhomogeneous earth. *Geophysics*, 48(8):1051–1062.

Jäger, R., Mann, J., Höcht, G., and Hubral, P. (2001). Common-reflection-surface stack: Image and attributes. *Geophysics*, 66(1):97–109.

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

Mann, J., Jäger, R., Müller, T., Höcht, G., and Hubral, P. (1999). Common-reflection-surface stack – a real data example. *Journal of Applied Geophysics*, 42(3-4):283–300.

Mayne, W. (1962). Common reflection point horizontal data stacking techniques. *Geophysics*, 27(6):927–938.

Netzeband, G., Gohl, K., Hübscher, C., Ben-Avraham, Z., Dehghani, G., Gajewski, D., and Liersch, P. (2006). The Levantine Basin – crustal structure and origin. *Tectonophysics*, 418(3-4):167–188.

Plessix, R. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.

Schwarz, B., Bauer, A., and Gajewski, D. (2016). Passive seismic source localization via common-reflection-surface attributes. *Studia Geophysica et Geodaetica*, 60(3):531–546.

Schwarz, B., Vanelle, C., Gajewski, D., and Kashtan, B. (2014). Curvatures and inhomogeneities: An improved common-reflection-surface approach. *Geophysics*, 79(5):S231–S240.

APPENDIX A

We will derive the adjoint-state method formula (24) for the gradient computing. Adjoint-state variables $\lambda_Q(x,t)$ and $\lambda_P(x,t)$ are some functions not defined so far. Let us take a dot product of equation (16) with vector $(\lambda_P, -\lambda_Q)$, integrate it by travel time from source to receiver and sum over all receivers

$$\overline{\triangle J} = \sum_{i=1}^{N_{data}} \int_0^{\tau(x_i)} dt \left(\lambda_P \left(\frac{\partial}{\partial t} \triangle\hat{Q} - v^2 \triangle\hat{P} - \triangle S_{11}\hat{Q} - \triangle S_{12}\hat{P} \right) \right.$$
$$\left. - \lambda_Q \left(\frac{\partial}{\partial t} \triangle\hat{P} + \frac{1}{v}\frac{\partial^2 v}{\partial q^2} \triangle\hat{Q} - \triangle S_{21}\hat{Q} - \triangle S_{22}\hat{P} \right) \right). \tag{29}$$

$\overline{\triangle J} = 0$ since the system is satisfied. We want to transform this expression to get perturbations $\triangle\hat{Q}, \triangle\hat{P}$ as multipliers. For this we integrate by parts

$$\overline{\triangle J} = \sum_{i=1}^{N_{data}} (\lambda_P \triangle\hat{Q})|_{t=0}^{t=\tau(x_i)} - \sum_{i=1}^{N_{data}} (\lambda_Q \triangle\hat{P})|_{t=0}^{t=\tau(x_i)}$$
$$- \sum_{i=1}^{N_{data}} \int_0^{\tau(x_i)} dt \left(\frac{\partial}{\partial t} \lambda_P \triangle\hat{Q} + \lambda_P v^2 \triangle\hat{P} + \lambda_P \triangle S_{11}\hat{Q} + \lambda_P \triangle S_{12}\hat{P} \right)$$
$$+ \sum_{i=1}^{N_{data}} \int_0^{\tau(x_i)} dt \left(\frac{\partial}{\partial t} \lambda_Q \triangle\hat{P} - \lambda_Q \frac{1}{v}\frac{\partial^2 v}{\partial q^2} \triangle\hat{Q} + \lambda_Q \triangle S_{21}\hat{Q} + \lambda_Q \triangle S_{22}\hat{P} \right). \tag{30}$$

As was previously mentioned $\triangle\hat{Q}(x, \tau(x)) = 0$ and $\triangle\hat{P}(x, \tau(x)) = 0$. To remove the external terms we impose initial conditions $\lambda_Q(x_i, 0) = 0$, $\lambda_P(x_i, 0) = 0$. The formula (15) should be modified to include the integration in time

$$\triangle J = \sum_{i=1}^{N_{data}} \hat{Q}(x_i, 0) \triangle\hat{Q}(x_i, 0) - \sum_{i=1}^{N_{data}} \hat{Q}(x_i, \tau(x_i)) \triangle\hat{Q}(x_i, \tau(x_i))$$

$$= -\sum_{i=1}^{N_{data}} \int_0^{\tau(x_i)} dt \frac{\partial}{\partial t}(\hat{Q}(x_i,t)\triangle\hat{Q}(x_i,t)) = -\sum_{i=1}^{N_{data}} \int_0^{\tau(x_i)} dt(v^2\hat{P}\triangle\hat{Q}+v^2\hat{Q}\triangle\hat{P}+\hat{Q}(\triangle S_{11}\hat{Q}+\triangle S_{12}\hat{P})). \tag{31}$$

$$\triangle J = \triangle J + \overline{\triangle J} \tag{32}$$

and the last step is to define the adjoint-state variables by forcing the sum of terms with $\triangle\hat{Q}$ and the sum of terms with $\triangle\hat{P}$ to be zero. This leads to the system of differential equations (25) and the formula (24) for the gradient computing.