Tomographic full waveform inversion and linear modeling of multiple scattering

Biondo Biondi

ABSTRACT

I present a tomographic full waveform inversion method that is based to an extension of the velocity model in time. The resulting wavefield modeling operator is linear with respect to the non-zero time lags of the extended velocity, but can effectively model multiple scattering caused by velocity perturbations. This property is attractive to achieve robust global convergence in a waveform inversion algorithm. A simple 1D numerical example illustrates the properties of the new modeling operator and its promises for robust waveform inversion.

INTRODUCTION

In a recent report Biondi and Almomin (2012) and Almomin and Biondi (2012) presented waveform-inversion methods with robust convergence characteristics even when the initial velocity model is far away from the correct one. These methods are based on an extension of velocity and reflectivity along the subsurface offset axes. This extension enables the kinematics of reflected to be correctly modeled by a linear operator even when the velocity errors are large. However, the extension also explodes the null space of the inverse problem. To ensure convergence towards desirable models a tomographic term is added to the inversion objective function that penalizes velocity models with energy at non-zero subsurface offsets.

In this paper I introduce a tomographic full waveform inversion (TFWI) that is based on an extension of the velocity model along the time axis instead of the subsurface offset axes. This time extension has the theoretical advantage that it can be directly linked to the modeling of multiple scattering phenomena; therefore, overcoming the limitations of conventional full waveform inversion (FWI), whose gradient is based on a first-order scattering approximation. Furthermore, the velocity extension along the time axis should enable robust convergence from transmitted, or refracted events, in addition to reflected events. This versatility can be beneficial when inverting long offset data that contain overturned and refracted events as well conventional reflections. A one-dimensional extension along time is also computationally more efficient than a two-dimensional extension along subsurface offsets. This is an important practical advantage since the computational cost of modeling wave propagation in extended velocity models is substantially larger than in conventional velocity models (Almomin, 2012).

Throughout this paper I illustrate the theory with simple 1D examples. Waves are propagated in 1D, and model parameter, both background and perturbations, are averaged over the whole propagation interval. In addition to be fast to compute by using Matlab, the 1D examples have the advantage of reducing the dimensionality of the model space and thus making the analysis of behavior of objective functions and gradients illustrative of the more general conceptual contributions of the paper. The numerical examples describes a transmission tomography problem to illustrate the capability of the proposed method to effectively use transmitted events, in addition to reflected ones.

CONVENTIONAL FULL WAVEFORM INVERSION (FWI)

Conventional full waveform inversion is performed by solving the following optimization problem

$$\min_{\mathbf{v}^2} J_{\mathrm{FWI}}\left(\mathbf{v}^2\right) \tag{1}$$

where:

$$J_{\mathrm{FWI}}\left(\mathbf{v}^2\right) = \frac{1}{2}\left\|\mathcal{L}\left(\mathbf{v}^2\right) - \mathbf{d}\right\|_2^2, \tag{2}$$

$\mathbf{v} = v\left(\vec{x}\right)$ is the velocity vector, $\mathcal{L}$ is a wave-equation operator non linear with respect to velocity perturbations and the data vector $\mathbf{d}$ is the pressure field $\mathbf{P} = P\left(t, \vec{x}\right)$ measured at the surface.

The wave-equation operator is evaluated by recursively solving the following finite difference equation

$$\left[\mathbf{D_2} - \mathbf{v}^2\nabla^2\right]\mathbf{P} = \mathbf{f}, \tag{3}$$

where $\mathbf{D_2}$ is a finite-difference representation of the second derivative in time, ∇^2 is a finite-difference representation of the Laplacian, and $\mathbf{f}$ is the source function.

Gradient computation with FWI

The efficient solution of the optimization problem expressed in equation 1 is performed by gradient based methods, and thus requires the evaluation of the linear operator $\mathbf{L}$, which is the linearization of $\mathcal{L}$ with respect to velocity perturbations $\delta\mathbf{v}^2$. This linear operator can be derived by perturbing equation 3 as follows

$$\left[\mathbf{D_2} - \left(\mathbf{v_o}^2 + \delta\mathbf{v}^2\right)\nabla^2\right]\left(\mathbf{P_o} + \delta\mathbf{P_o}\right) = \mathbf{f}, \tag{4}$$

where $\mathbf{P_o}$ and $\mathbf{v_o}$ are the background wavefield and velocity, respectively, and $\delta\mathbf{P_o}$ is the scattered wavefield.

Equation 4 can be rewritten as the following two equations:

$$\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\mathbf{P_o} = \mathbf{f}, \tag{5}$$

$$\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\delta\mathbf{P_o} = \delta\mathbf{v}^2\nabla^2\left(\mathbf{P_o} + \delta\mathbf{P_o}\right), \tag{6}$$

which represents a nonlinear relationship between velocity perturbations and scattered wavefield. To linearize this relationship we drop the term multiplying the perturbations with each other; that is, we drop the scattered wavefield from the right hand side of equation 6 and obtain the following coupled equations:

$$\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\mathbf{P_o} = \mathbf{f}, \tag{7}$$

$$\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\delta\mathbf{P_o} = \delta\mathbf{v}^2\nabla^2\mathbf{P_o}. \tag{8}$$

The linear operator $\mathbf{L}$ used to compute the gradient of the FWI objective function 2 is evaluated by recursively propagating the background wavefield $\mathbf{P_o}$ and the scattered wavefield $\delta\mathbf{P_o}$ by solving equations 7–8.

The scattered wavefield $\delta\mathbf{P_o}$ is a linear function of the velocity perturbations $\delta\mathbf{v}^2$ because equation 8 takes into account only fist order scattering. Notice that the linear operator $\mathbf{L}\left(\mathbf{v_o}^2\right)$ is itself a non linear function of the background velocity, both directly by determining the propagation speed of the scattered wavefield (left hand side in equation 8), and indirectly through the background wavefield (right hand side in equation 8).

Problems with FWI

Unfortunately, high-order scattering must be taken into account to model accurately wavefield perturbations when the velocity perturbations have wide spatial extent and/or large amplitude. Such velocity perturbations cause significant (larger than one fourth of wave cycle) time shifts in the propagating wavefield. The linear operator $\mathbf{L}$ cannot model large time shifts because the source function on the right-hand side of equation 8 is triggered by the background wavefield reaching a velocity perturbation, and consequently it has the timing as the background wavefield. Furthermore, the perturbed wavefield is propagated with the background velocity $\mathbf{v_o}$. In mathematical terms

$$\mathcal{L}\left(\mathbf{v_o}^2 + \delta\mathbf{v}^2\right) \neq \mathcal{L}\left(\mathbf{v_o}^2\right) + \mathbf{L}\left(\mathbf{v_o}^2\right)\delta\mathbf{v}^2. \tag{9}$$

The problem is even deeper. When $\delta\mathbf{v}^2$ causes large time shifts by multiple scattering, there is no perturbation $\widehat{\delta\mathbf{v}^2}$ that can model those time shifts by single scattering; that is,

$$\mathcal{L}\left(\mathbf{v_o}^2 + \delta\mathbf{v}^2\right) \neq \mathcal{L}\left(\mathbf{v_o}^2\right) + \mathbf{L}\left(\mathbf{v_o}^2\right)\widehat{\delta\mathbf{v}^2} \text{ for any } \widehat{\delta\mathbf{v}^2}. \tag{10}$$

The non linearity of the modeling operator makes the objective function equation 2 to be non convex when the velocity perturbations are sufficiently large. Figure 1 shows an example of non-convexity of the objective function. The result correspond to several 1D transmission problems sharing the same starting velocity (1.2 km/s) and with different true velocities. For all these experiments the source-receiver offset is 4 km and the source function is a zero-phase wavelet bandlimited between 5 and 20 Hz. The FWI norm is plotted as a function of the true velocity. If the true velocity is

lower than ≈ 1.18 km/s or larger than ≈ 1.22 km/s a gradient based method will not converge to the right solution, even in this simple and low-dimensionality example.

The challenges of solving the optimization problem in equation 1 by gradient based optimization can be alternatively represented by graphing, as a function of the initial velocity error, the search direction (opposite sign of the gradient direction) of the objective function with respect to velocity square. Figure 2 display this function computed by applying the adjoint of the linear operator $\mathbf{L}$ to the data residuals; that is

$$\nabla J_{\mathrm{FWI}} = \mathbf{L}' \left[\mathcal{L}\left({\mathbf{v_o}}^2 \right) - \mathbf{d} \right]. \tag{11}$$

For a gradient based method to converge, the search direction should be always negative when the true velocity is lower than 1.2 km/s, and positive when the true velocity is higher.

Multiple-scattering modeling

We can achieve accurate modeling of perturbed wavefield by solving equations 5–6 instead of equations 7–8. Equations 5–6 can be solved numerically with a simple explicit method; that is, one that adds the scattered wavefield up to time t to the right-hand side of equation 6 to compute the scattered wavefield at $t + \Delta t$. Even in presence of large velocity variations, the scattered wavefield has now the correct time shift. Numerical solutions produce accurate results, although the scattered wavefield is still propagated with the background velocity, because multiple scattering is taken into account of.

The challenge with using these equations in a gradient-based inversion algorithm is that the relation between the scattered wavefield $\delta\mathbf{P_o}$ and the velocity perturbations $\delta\mathbf{v}^2$ is now nonlinear. In the next section, I present a method for linearizing this relation that is alternative to the conventional one represented by equations 7–8, and is based on an extension of the velocity model in time.

TOMOGRAPHIC FULL WAVEFORM INVERSION (TFWI)

We can rewrite equations 5–6 by performing the following substitution

$$\delta\mathbf{P_o} = \mathbf{P_o} \overset{t'}{*} (\mathbf{T} - \mathbf{I}), \tag{12}$$

and consequently

$$\mathbf{P_o} + \delta\mathbf{P_o} = \mathbf{P_o} \overset{t'}{*} \mathbf{T}, \tag{13}$$

where $\mathbf{T}$ is a convolutional operator in time that may vary both in space and time; $\mathbf{I}$ is the identity operator. For example, when the perturbed wavefield is a time-shifted

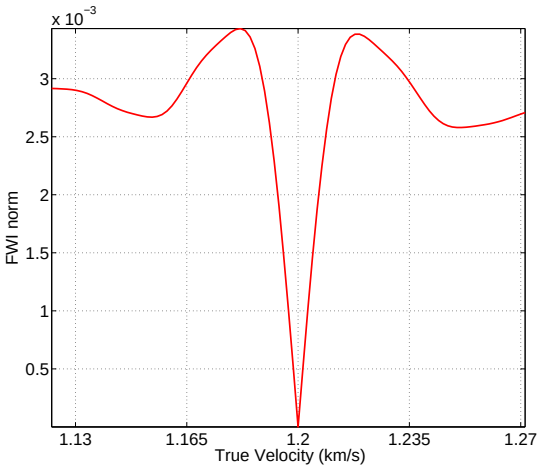

Figure 1: FWI norm as a function of the true velocity, when the starting velocity is equal to 1.2 km/s.

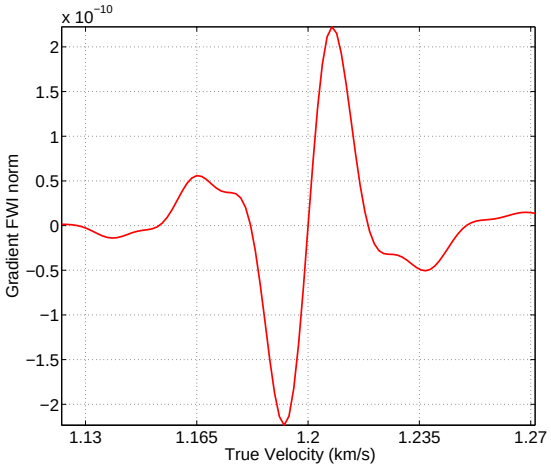

Figure 2: FWI search direction as a function of the true velocity, when the starting velocity is equal to 1.2 km/s.

version of the background wavefield, the operator $\mathbf{T}$ is a shifted delta function. With this substitution equation 6 can be rewritten as

$$\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\delta\mathbf{P_o} = \delta\mathbf{v}^2\left(\widetilde{\mathbf{T}} \overset{t'}{*} \nabla^2\mathbf{P_o}\right), \tag{14}$$

where the substitution of $\mathbf{T}$ with $\widetilde{\mathbf{T}}$ takes into account of the Laplacian.

If we define an velocity model *extended* in time $\delta\tilde{\mathbf{v}}^2\left(t,t'\right) = \delta\mathbf{v}^2 \overset{t'}{*} \widetilde{\mathbf{T}}$, we can rewrite equation 14 as

$$\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\delta\mathbf{P_o} = \delta\tilde{\mathbf{v}}^2\left(t,t'\right) \overset{t'}{*} \nabla^2\mathbf{P_o}. \tag{15}$$

The estimation of an extended velocity as a function of both t and t', and for each seismic experiment (e.g. shot), can be unpractical. We can approximate equation 15 by making the velocity dependent only from the convolutional time lag; that is, $\tau = t - t'$ and the same for each seismic experiment. The approximation of equation 15 can be written as

$$\left[\mathbf{D_2} - \tilde{\mathbf{v}}^2\left(\tau=0\right)\nabla^2\right]\mathbf{\Delta P} = \delta\tilde{\mathbf{v}}^2\left(\tau\right) \overset{\tau}{*} \nabla^2\mathbf{P_o}, \tag{16}$$

where the change of notation from $\delta\mathbf{P_o}$ to $\mathbf{\Delta P}$ indicates that the scattered wavefield $\mathbf{\Delta P}$ is now an approximation of the true multiple-scattered wavefield $\delta\mathbf{P_o}$.

Formally solving equation 16 we obtain

$$\mathbf{\Delta P} = \left[\mathbf{D_2} - \tilde{\mathbf{v}}^2\left(\tau=0\right)\nabla^2\right]^{-1}\left[\delta\tilde{\mathbf{v}}^2\left(\tau\right) \overset{\tau}{*} \nabla^2\mathbf{P_o}\right] \tag{17}$$

that is a linear relationship between $\delta\tilde{\mathbf{v}}^2$ and $\mathbf{\Delta P}$ defined by the linear operator $\widehat{\mathbf{L}}$ such as $\mathbf{\Delta P} = \widehat{\mathbf{L}}\delta\tilde{\mathbf{v}}^2$.

If we define the total wavefield to be

$$\mathbf{P} = \mathbf{P_o} + \mathbf{\Delta P}, \tag{18}$$

and the extended non-linear modeling operator as

$$\tilde{\mathcal{L}}\left(\tilde{\mathbf{v}}\right) = \mathcal{L}\left(\tilde{\mathbf{v}}\left(\tau=0\right)\right) + \widehat{\mathbf{L}}\left(\tilde{\mathbf{v}}\right)\delta\tilde{\mathbf{v}}^2, \tag{19}$$

the objective function

$$J_{\mathrm{EFWI}}\left(\tilde{\mathbf{v}}\right) = \frac{1}{2}\left\|\tilde{\mathcal{L}}\left(\tilde{\mathbf{v}}\right) - \mathbf{d}\right\|_2^2 \tag{20}$$

has the same local minima of the original FWI objective function, but it also provides smooth descending paths to the global minimum in the additional dimensions. The problem is now under constrained because many solutions fit the data equally well. Among all these possible solutions we are interested in the solutions for which the extended velocity model is as focused as possible around the zero time lag of the model.

To converge towards a desirable solution we can add an additional term to the objective function that penalizes extended velocity model with significant energy at non-zero time lag; that is,

$$\min_{\tilde{\mathbf{v}}} J_{\mathrm{TFWI}}\left(\tilde{\mathbf{v}}\right), \tag{21}$$

with

$$J_{\mathrm{TFWI}}\left(\tilde{\mathbf{v}}\right) = \frac{1}{2}\left\|\tilde{\mathcal{L}}\left(\tilde{\mathbf{v}}\right) - \mathbf{d}\right\|_2^2 - \epsilon\mathcal{F}\left(\tilde{\mathbf{v}}\right), \tag{22}$$

where $\mathcal{F}$ is an operator that measure the focusing of the model at zero time lag. A straightforward example of such operator is

$$\mathcal{F}\left(\tilde{\mathbf{v}}\right) = -\left\||\tau|\,\tilde{\mathbf{v}}\right\|_2^2. \tag{23}$$

Gradient computation with TFWI

The gradient computation of the TFWI objective function has three components. The first component takes into account the dependency of the background wavefield from velocity; it is the same as for the FWI case. The second component is related to the dependency of the approximation of the scattered wavefield $\mathbf{\Delta P}$ from the extended velocity. It is performed by applying the adjoint of the linear operator $\widehat{\mathbf{L}}$. The application of $\widehat{\mathbf{L}}'$ to the data residual is accomplished similarly to the FWI gradient by injecting the residual at the receiver location and running backward in time the propagation expressed in equation 17.

The third component takes into account the dependency of $\mathbf{\Delta P}$ from the velocity at zero time lag; its evaluation is more involved than for the previous two terms but is crucial to the convergence of the inversion towards a velocity model that explain the kinematics in the data. The forward operator can be evaluated as a chain of two operators. The first one relates perturbations in velocity to perturbations in the background wavefield, as expressed by equations 7–8. The second one is computed by forward solving in time equation 17, where the source term is given by the perturbations in the background wavefield, and not by the perturbations in the extended velocity. Consequently this term is zero when the extended velocity perturbations $\delta\tilde{\mathbf{v}}^2$ are zero, independently from the perturbations in the background wavefield. The adjoint is computed by applying the adjoint of these two operators in reverse order.

Finally, the gradient of the regularization terms depends on the expression of the specific focusing operator $\mathcal{F}$. For the choice expressed in equation 23, the computation of the gradient is trivial.

NUMERICAL 1D EXAMPLE

I will use a simple 1D numerical example to analyze some of the characteristics of the TFWI method I presented in the previous sections. Figure 3 shows the difference between the background wavefield propagated with $\mathbf{v}$=1.2 km/s and the wavefield

propagated with the true velocity of $\mathbf{v}$=1.13 km/s. The difference wavefield is displayed as a function of propagation distance and traveltime. The data are recorded with a receiver located at 7 km for a total of 4 km offset from the source. The velocity error is sufficiently high that the wavefields are completely out of phase at the receiver location. We are therefore in the situation described by equation 10. Because the events are out of phase at the receiver, the backprojection of the data residual into the velocity model yields a gradient (equation 11) that is substantially zero, as it can be verified in Figure 2. Conventional FWI would have troubles to converge even for this simple problem.

On the contrary, the linearized modeling equation defined in equation 17 would have no troubles to model the data residual. For example, we can easily reproduce the wavefield difference shown in Figure 3 by setting the extended-velocity perturbation to be a delta function along the τ axis, where the shift of the delta function linearly increases with the distance from the origin. This linear shift is computed by integrating the difference in slowness between the background model and the true model. The extended-velocity perturbation is shown in Figure 4. Figure 5 shows the result of solving equation 17 with the model shown in Figure 4. The approximation of the scattered wavefield $\mathbf{\Delta P}$ is almost identical to the wavefield difference shown in Figure 3.

Next step is the backprojection of the data residual recorded by extracting the wavefields at the receiver position into the extended velocity model. In the previous section, I explained that there are three terms in this backprojection. In this case, the first and the third are zero. As discussed previously, the first component is zero because the events in the data residual are out of phase. The third term is zero because the starting extended model is zero at non-zero time lags. Figure 6 shows the contribution of the second component of the gradient. It is zero on the right of the receiver location, and it is basically constant as function of the propagation distance on the left of the receiver location.

Forward modeling solving equation 17 with the model shown in Figure 6 yields a scattered wavefield that at the receiver location is extremely close to the original residual shown in Figure 3. Therefore, the first term of the objective function 22 has a well-behaved parabolic shape as a function of the step size applied to the search direction, with well-defined minimum that determines the extended model after a first iteration of an iterative inversion algorithm.

The extended model obtained at the first iteration can then be used to compute the data residuals and gradient at the second iteration. Since the data residual are small, the main contribution to the gradient comes from the second term of the objective function; that is, the focusing operator. It is straightforward to verify that for the choice of focusing operator in 23, the most significant component of this gradient is away from the zero time lag. The projection of this gradient in the data space has a non-zero component and will create a data residual, which can in turn be backprojected into the velocity model. The most interesting component of this backprojection is at zero-time lag since it is the one that will effect the propagation

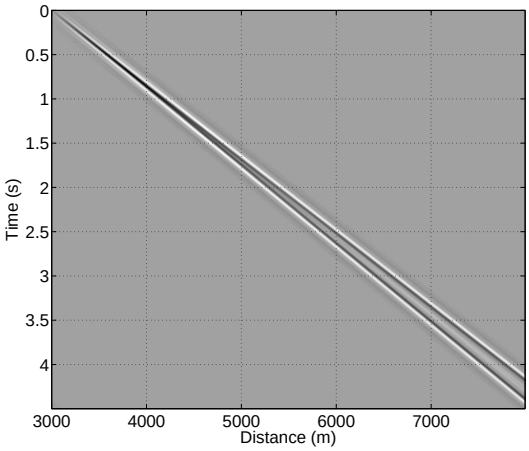

Figure 3: Difference between background wavefield computed with the starting velocity (1.2 km/s) and the wavefield propagated with the true velocity (1.13 km/s).

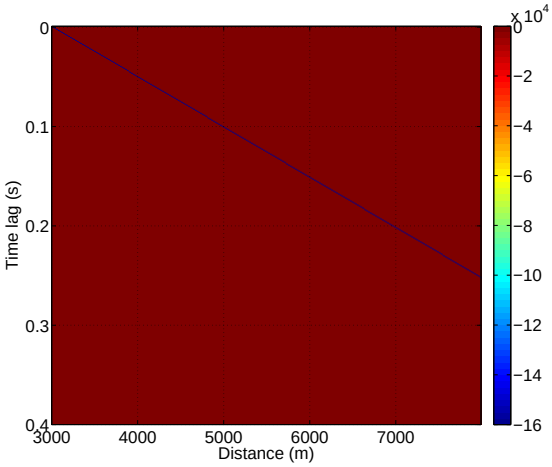

Figure 4: Extended velocity perturbation chosen to approximately model the wavefield difference shown in Figure 3 by applying equation 17.

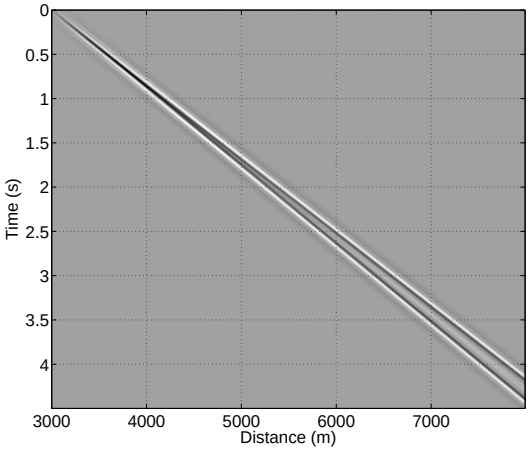

Figure 5: Perturbed wavefield computed by solving equation 17 with the model shown in Figure 4.

velocity of the background wavefield at the next iteration.

Among the three gradient components discussed in the previous section, only the third one effects the extended velocity at zero-time lag. Figure 7 shows the search direction obtained by averaging this gradient component along the whole propagation interval as a function of the true velocity. As before the the starting velocity is 1.2 km/s. In contrast with the conventional FWI search direction shown in Figure 2, the search directions shown in Figure 7 is always negative for a true velocity lower than 1.2 km/s, and always positive for a true velocity larger than 1.2 km/s. This indicates that no matter how large the initial velocity error, the TFWI method will start moving the zero-lag component of the extended model in the correct direction. This result is far from being a proof of global convergence, but is definitely encouraging.

CONCLUSIONS

Perturbations in the propagating wavefield caused by multiple scattering can be approximately modeled by a linear operator when the velocity function is extended in time. A tomographic full waveform method based on a time extension of the velocity model is likely to have attractive global-convergence characteristics, and thus to overcome one of the main challenges of conventional full waveform inversion.

REFERENCES

Almomin, A., 2012, Computational analysis of extended full waveform inversion: SEP-Report, **148**, 59–70.

Almomin, A. and B. Biondi, 2012, Tomographic full waveform inversion: Practical and computationally feasible approach: SEP-Report, **147**, 13–26.

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: SEP-Report, **147**, 1–12.

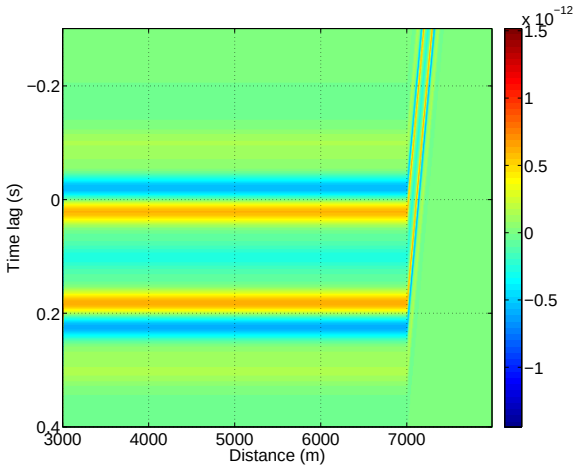

Figure 6: Extended-velocity gradient computed by applying the adjoint of the linear operator $\widehat{\mathbf{L}}$ to the data residuals measured at the receiver located at a distance of 7 km.

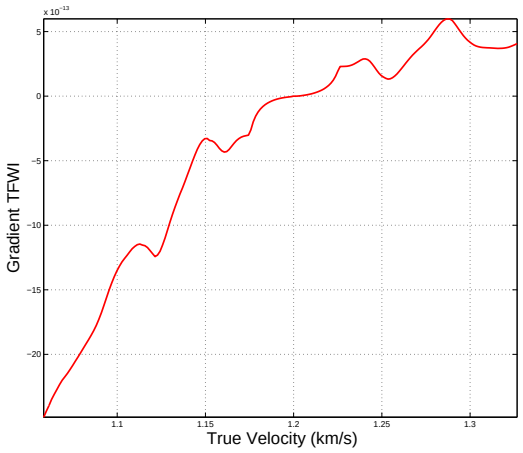

Figure 7: Search directions obtained at the second iteration of the inversion process and caused by the focusing term of the TFWI objective function. These search directions are obtained by averaging the gradient along the whole propagation interval. Notice that the search directions are always negative for a true velocity lower than 1.2 km/s, and always positive for a true velocity larger than 1.2 km/s.