EFFICIENT DEFLECTION ANGLE BASED FILTERING FOR FWI

V. Kazei and E. Tessmer

email: *vkazei@gmail.com*

keywords: *scattering angle, deflection angle, reverse-time migration, efficient filtering, full-waveform inversion,*

ABSTRACT

The importance to access the scattering angle information has been recognized in ray based applications a long time ago, but only recently became available in the field of wave equation based imaging and inversion. First it was implemented for wave equation migration velocity analysis then reverse-time migration and finally full-waveform inversion. Conventional access to the scattering angle information in seismic imaging via wavefields continuation requires an extension either in space or in time which is costly in terms of computational resources. For a single frequency this filtering can however be interpreted as a non-stationary convolutional filtering. This works quite well but still not efficient enough for 3D models. For the purpose of a more efficient scattering angle filter, we develop techniques that utilize the mapping nature (no domain extension) of the scattering angle based filter for constant-velocity background models. We split the background velocity model into regions with different velocity ranges, generating an "extension in velocity", so that in each region the velocity is assumed not to vary much. We establish an analytical relation between this new extension and conventional time lag extension. By numerical example we show that a few samples in the newly introduced dimension is enough to construct the efficient scattering angle filter. The filter can be utilized either for full-waveform inversion preconditioning or to clean up reverse-time migration artifacts. A new way for the interpolation by splitting the background velocity model with a smooth decomposition of unity is also proposed.

INTRODUCTION

Conventional full-waveform inversion (FWI) faces convergence problems in case of the lack of low frequencies acquired. These problems are usually treated by replacing the misfit functional minimized (Van Leeuwen and Mulder, 2010; Bozdağ et al., 2011; Wu et al., 2014). The other way to tackle the problem is expanding the search space (van Leeuwen and Herrmann, 2013). Advanced multidimensional optimization techniques like the L-BFGS method (Nocedal, 1980) are utilizing only the information contained in the misfit gradients, so the other option is to apply conditioning of the full-waveform inversion gradients. Since it is well known that in the presence of low frequencies in the data FWI converges well, we design a filter to mimic the gradient behavior at lower frequencies by conditioning the gradients at high frequencies with a filter. The linear component of waveform inversion, the gradient, reflected in an imaging-like process (Tarantola, 1984), involves correlation of two wavefields. In the classical full-waveform inversion (FWI) implementation, one wavefield corresponds to the source, referred to as the state variable, and the other corresponds to the receiver wavefield for that source, referred to as the adjoint variable (Sirgue and Pratt, 2004). The resolution or the smoothness of the resulting gradient, governed by diffraction tomography principles (Devaney, 1984; Mora, 1989; Kazei et al., 2013), is controlled by the angle between the two interacting wavefields (Figure 1) as well as by their local wavelengths λ in the following way:

- if the media is well illuminated by transmissions the local low wavenumbers can be resolved by FWI everywhere since little deflection angles are accessible for most of its parts
- if high frequencies are present in the data, they are overwhelming the low wavenumber content of the gradient and mask it due to much higher amplitudes (Sirgue and Pratt, 2004),
- the local resolution of FWI is limited by a half of the shortest wavelength taken into consideration for the gradient evaluation (Virieux and Operto, 2009).

The local wavelengths are the same and equal to λ unless background models with converted waves attenuation or anisotropy are considered. The basis for this local plane wave analysis is given by the idea of Ewald's spheres construction (Ewald, 1969). The main outcome from this theory that we are going to use is the simple geometrical relation between the wavenumbers in a local perturbation and the scattering plane wave are related by the equation (1) which is valid within first order Born approximation.

$$\mathbf{K} = \frac{\pi}{\lambda(\mathbf{x})}(\mathbf{s} + \mathbf{g}). \tag{1}$$

For completeness the equation (1) is discussed and derived in Appendix A: Here $\mathbf{K}$ is the spatial wavenumber vector of the velocity perturbation that causes the particular scattering event, $\mathbf{s}$ and $\mathbf{g}$ are unit vectors towards source and receiver respectively. Since at each iteration in full-waveform inversion the Born approximation is used, the other interpretation of equation (1) is the relation between the wavenumber $\mathbf{K}$ updated and the incident directions of wavefields ($\mathbf{s}$,$\mathbf{g}$) that are correlated to obtain the gradient. Taking the absolute values on both sides in equation (1), we obtain the scalar version:

$$|\mathbf{K}| = \frac{2\pi}{\lambda(\mathbf{x})} \cos\frac{\theta}{2}. \tag{2}$$

As the process of imaging is meant to establish a form of continuation of the incident wavefield into

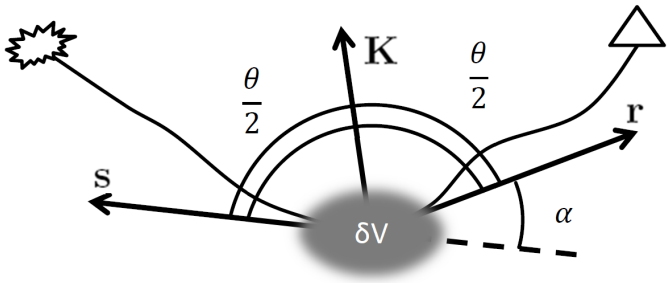


Figure 1: The scattering of a monochromatic locally plane wave by a perturbation of the background model.

the scattered wavefields, whether such continuation is based on a reflection point, or a mere transmission, the scattering angle defines the nature of the wavefield continuation. Since transmissions are a result of wavefields continuing in their path (like direct and diving waves), the scattering angle between the wavefield is approaching π in high-frequency asymptotics. On the other hand, reflections admit much smaller scattering angles, being near zero for the majority of zero offset energy. Thus the nature of waves contributing to an image or a gradient can be controlled by applying constraints to the scattering angles. It is well known that reflections close to backscattering are more suitable for migration while full-waveform inversion works better for transmissions at first iterations, so we can extract them in the image domain by keeping only appropriate scattering angles.

Imaging using models with sharp contrasts often yields low frequency noise in the image. Khalil et al. (2013) recognized that one can filter the image to remove the unwanted low frequency noise by using the scattering angle information, which leads to better images than conventional smoothing. They suggested an extended image using a normalized time lag to filter out the high-scattering angles related part

of the gradient. One can also filter the gradient to admit more smooth components to the velocity model. Alkhalifah (2015a) used this approach to develop an FWI gradient filtering process that tends to filter out the low scattering angles at the early stages of FWI. Also the angles between the two wavefields can easily be utilized for background velocity quality control purposes. Accessing such information required extension of the domain of imaging using space or time lags (Sava and Fomel, 2003).

The process of extending the gradient with an additional dimension can be overwhelming, especially in 3D seismic imaging. The stretched extension in time lag is needed to properly handle the velocity dependency nature of the non-stationary smoothing or sharpening operation given by the conventional scattering angle filter.

Here we propose two alternatives to the conventional scattering angle filter - the non-stationary Gaussian filtering (Margrave, 1998) and a replacement of time or space extension needed for scattering angle filter with splitting the background model into domains with different velocity ranges "extension in velocity". This allows us to mitigate extensions in space or time and perform the filtering much more efficiently.

Though an efficient implementation of the filtering can be used for mitigating low frequency artifacts and improving the image resolution, we will focus here on its use to filter the FWI gradient. The scattering angle filter can be applied directly to any adjoint state method based gradients. In the examples below, we use gradients obtained from conventional FWI.

Deflection, scattering and reflection angles

In Figure 1 we show deflection, scattering and reflection angles. θ is the scattering angle as it is usually introduced in geophysics, meanwhile $\theta/2$ is the angle of reflection. α is the scattering angle in many other areas of physics which consider scattering of particles (e.g. atomic physics, quantum mechanics or physics of elementary particles). At former times seismic measurements were concentrated on small offsets ($\theta \simeq 0$), thus the difference from other definition was reasonable for geophysics. With the development of seismic acquisition techniques as well as the processing algorithms such as full-waveform inversion, the long offsets and transmissions (diving waves $\alpha \simeq 0$) came into surface seismic. For FWI particularly at early stages the diving waves are crucial while the reflections are complicated to be incorporated into inversion. Therefore the scattering angle as it is more common for physics, not for geophysics becomes more reasonable to be used. Originally in the work by Rutherford (1911) α was called the angle of "deflexion". Since we are going to filter for transmissions ($\theta \simeq \pi$) we find it more reasonable to work with α. To remove the existing ambiguity between physics and geophysics we will follow Rutherford (1911) and call α the angle of deflection.

THE NON-STATIONARY FILTERING FOR SCATTERING ANGLE INFORMATION

For a constant velocity background, the filter is stationary and no extension is needed to apply the filter, since the local wavelength $\lambda(\mathbf{x})$ is constant restricting the scattering angles involved in imaging process is as simple as restricting the absolute values of updated wavenumbers in the model due to the equation (2):

$$\cos\frac{\theta}{2} = \frac{1}{2\pi}|\mathbf{K}|\lambda, \tag{3}$$

Thus one can easily construct the scattering angle based filter for constant velocity background:

$$\tilde{I}(\mathbf{x}) = F^{-1}w(\mathbf{K}, v, \omega)FI(\mathbf{x}), \tag{4}$$

where F and F^{-1} stands for the Fourier and inverse Fourier transform in space, respectively. To mitigate low frequency imaging artifacts Khalil et al. (2013) apply:

$$w(\mathbf{K}) = \begin{cases} 1, \text{ if } \cos\frac{\theta}{2} > \frac{v|\mathbf{K}|}{\omega}, \\ 0, \text{ if } \cos\frac{\theta}{2} \leq \frac{v|\mathbf{K}|}{\omega} \end{cases} \tag{5}$$

for a given angle θ.

To mitigate the nonlinear parts of the FWI gradient Alkhalifah (2015a) applies:

$$w(\mathbf{K}) = \begin{cases} 0, \text{ if } \cos\frac{\theta}{2} > \frac{v|\mathbf{K}|}{\omega}, \\ 1, \text{ if } \cos\frac{\theta}{2} \leq \frac{v|\mathbf{K}|}{\omega}. \end{cases} \tag{6}$$

In both cases (RTM and FWI), the filtering will benefit from a smoothing of the tapering function to reduce Gibbs phenomenon effect. For the purposes of smoothing of FWI gradients the Heaviside filter in equation (5) can be replaced by a Gaussian in the Fourier domain providing the same standard deviation (same second moment). This will result into a Gaussian filter in the spatial domain. In Appendix B we derive the parameters of the deflection angle based Gaussian filter which does "frequency scaling" for us and show how it is related to the scattering angle based filter introduced by Khalil et al. (2013). Gaussian filtering can be done using the local background velocity instead of the global average one (Ravaut et al., 2004). The Gaussian filtering has certain benefits, e.g. it lacks the Gibbs phenomenon and it has the least L_2 norms product of filters in space and in wavenumber spectrum. The last means that it will keep the filter as local as possible for a given spatial wavenumber range:

$$\tilde{I}(\mathbf{x}) = \int\int \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{\lambda^2(\mathbf{x})}\right) I(\mathbf{x}')\mathrm{d}\mathbf{x}, \tag{7}$$

where the local length of the filter is

$$\lambda(\mathbf{x}) = \frac{2\pi v(\mathbf{x})}{\omega_{filt}}, \text{ where } \omega_{filt} = \omega\cos\frac{\theta}{2}. \tag{8}$$

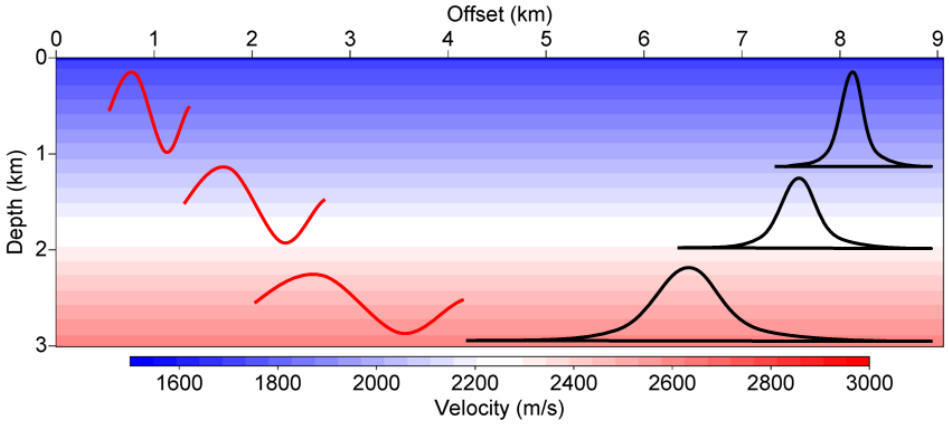
The length of the filter has to vary to compensate for non-homogeneous background media. ω_{filt} is an effective frequency of the gradient after filtering. In a way the wavenumber content is scaled with $\cos\frac{\theta}{2}$. Margrave (1998) calls this type of the filtering the non-stationary combination, however we refer to all similar filters as to a non-stationary convolution since from the formal mathematical point of view there is no difference.

Inversion for the Marmousi model

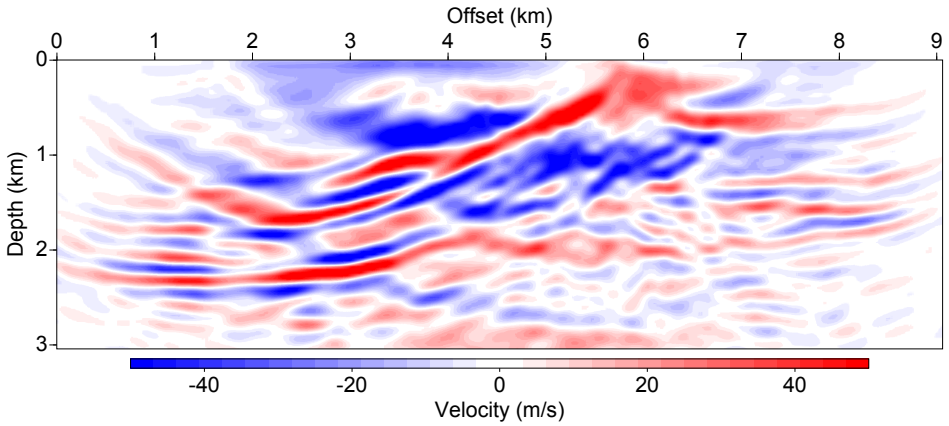
We show the application of the frequency scaling filter on the example of inversion for the Marmousi model (Bourgeois et al., 1991). For modeling we use a pseudo-spectral method (Kosloff and Baysal, 1982; Tessmer, 2011; Virieux et al., 2011) with 4-th order Taylor expansion in time for time stepping. Pseudo-spectral modeling allows us to keep the model very sparse, having 2 grid points per shortest wavelength, which matches in the low velocity regions the highest FWI resolution of half of the local wavelength (Virieux and Operto, 2009). We start from the model with velocity linearly increasing in depth (Figure 2(a)). For the Marmousi example we choose the grid step equal to 32 m. At all sides of the model we attach 1 km wide sponge absorbing boundaries. The synthetic data from 4 shots were acquired at 276 receiver locations at the surface. First, we try to use the conventional multiscale full-waveform inversion (Bunks et al., 1995) of 5 Hz central frequency data. From the very beginning all the parts of the model with high frequency content were updated as follows from the initial gradient (Figure 2(b)) and that probably leads to the convergence problems and fails as shown in Figure 4(a).

Trying to improve the results with conventional Gaussian filtering with length of the shortest wavelength also fails. Convergence can however be achieved by filtering the gradients with the frequency scaling non-stationary filter with length proportional to the local wavelength at a lower frequency than available from the data (Figure 2(a)). The main feature of the non-stationary filtering approach is that it provides a very detailed image of the upper part while the deeper is not updated much due to stronger gradient smearing compared to the regular Gaussian filtering (Figure 3(a)). This feature is related only to background models that have velocity increasing with depth. In general this scattering angle based filtering can be used for scale separation in algorithms combining wave equation migration velocity analysis (WEMVA) and FWI, e.g. reflection full-waveform inversion RFWI (Alkhalifah, 2015a) or RWI (Zhou et al., 2015) or Tomographic Full-Waveform Inversion (TFWI) (Biondi and Almomin, 2013). Inversion of the same data regularized by the frequency scaling filter converges Figure 4(b).

Thus the benefits of this non-stationary scattering angle filter for full-waveform inversion convergence are almost the same as for the conventional one (Alkhalifah, 2015b).

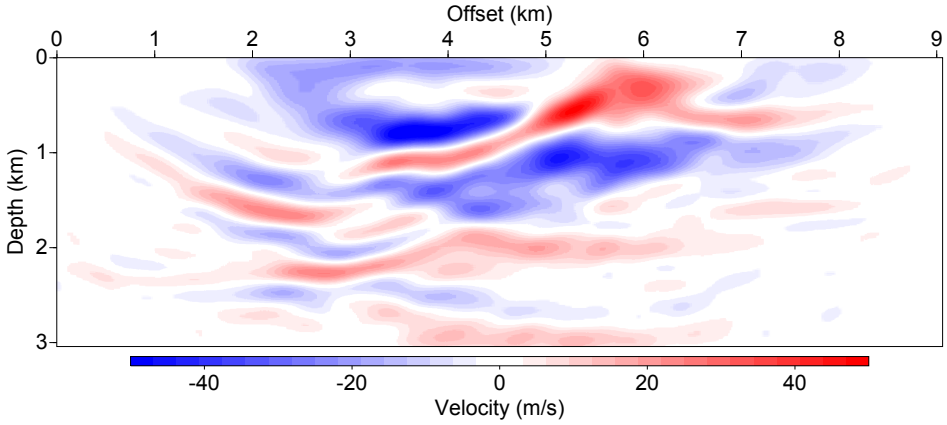


(a)

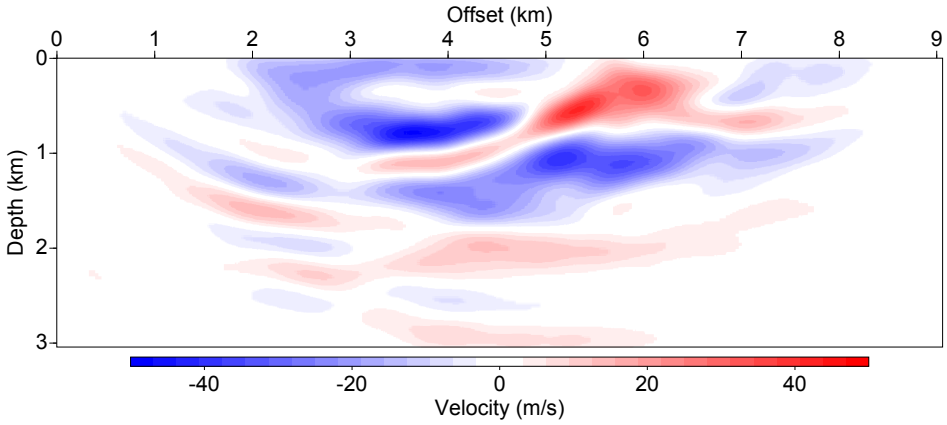


(b)

Figure 2: Filter within the starting model for inversion for Marmousi. The length of the filter (black) is governed by the local wavelength (red) at a desired, lower frequency than available from the data. One of the most important features of the filter is that it is sharper at the low velocity regions (which are usually close to the surface) than at the deeper high velocity parts, retaining this important property of the gradients constructed from the low frequency data (a). Full-waveform inversion gradient at first iteration when inverting the signal 4-6 Hz, using the initial model (b).



(a)



(b)

Figure 3: The gradient filtered with the stationary Gaussian filter of length equal to the shortest wavelength in the model at 2 Hz (a). Gradient filtered with the scattering angle based non-stationary gaussian filtering (b). It is visible that in the upper part both filters provide almost the same results, while in the deeper part the scattering angle based filtering reduces the resolution much more significantly. All the images are scaled with depth to make changes in the deeper parts visible.

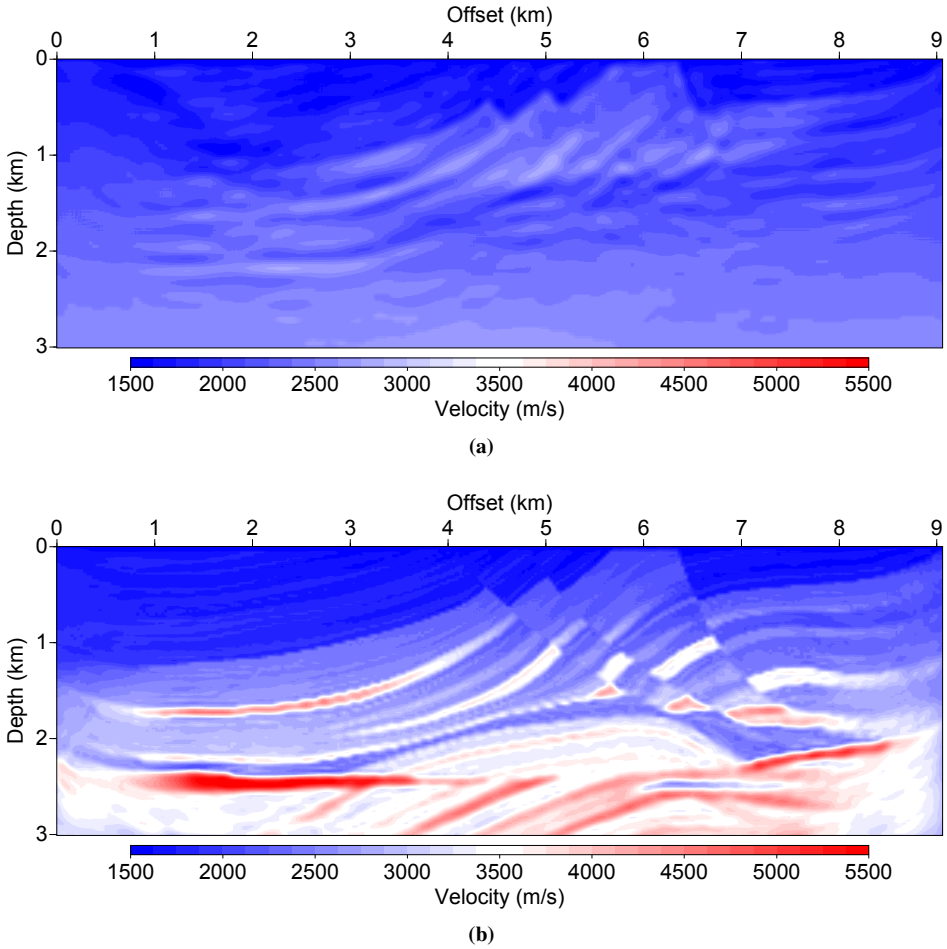


Figure 4: Inversion results for 4-6 Hz data: no filters applied (a), inversion with the scattering angle filter applied to mimic the lacking frequency of 2.5 Hz before conventional multiscale full-waveform inversion (b).

EFFICIENT FILTERING BY EXTENSION IN THE VELOCITY

Conventional extension

Using the normalized time-lag representation (Khalil et al., 2013), the process of a gradient or an image extension in the frequency domain is given by the following formula

$$I(\mathbf{x},\zeta) = I(\mathbf{x})e^{-2i\omega\frac{\zeta}{v(\mathbf{x})}}, \tag{9}$$

where $I(\mathbf{x})$ is the image and v is the velocity defined using the space variable vector $\mathbf{x}$, ζ is the time normalized extension parameter, and ω is the angular frequency. An inherent feature of this modified time-lag (distance units) representation is that the relationship of the scattering angle to the wavenumber of the gradient is free of a velocity (space) dependency. In fact, the scattering angle θ is then given by the following formula:

$$\cos^2\frac{\theta}{2} = \frac{|\mathbf{K}|^2}{k_\zeta^2}, \tag{10}$$

where $\mathbf{K}$ is the wavenumber vector and k_ζ is the wavenumber (Fourier transform) corresponding to ζ. A four-dimensional Fourier transform of $I(\mathbf{x},\zeta)$ (three-dimensional in 2D), will allow us to map $I(\mathbf{K},k_\zeta)$ to its angle gather equivalent $I(\mathbf{K},\theta)$, using equation 10. Extending an image in time for the conventional scattering angle based filtering (Sava and Fomel, 2003) can be computationally overwhelming especially in 3-D seismic imaging. To make it more efficient we consider a monochromatic gradient filtering for a single pair of source and receiver. The conventional gradient obtained for ray-Born inversion (Beydoun and Mendes, 1989) is shown at the Figure 6(a). Using a scattering angle filter that allows only angles above 170 degrees, we end up with the gradient shown in Figure 6(b). The filtering required the computation of an extended axis for ζ, as shown in Figure 7. This adds a cost of order $N\log N$ to a direct wavenumber filtering to the gradient, where N is the number of samples of the extended axis ζ. In this example, we used 200 ζ samples at a sampling rate of 0.02 km, which also increased our storage requirements 200 times. We would like to outline that consideration of this eikonal based gradient allows us to isolate the problem of geometrical spreading compensation, which is treated by quasi-Newton methods within regular full-waveform inversion. Following the equation (9) we generate the extension shown in Figure 7.

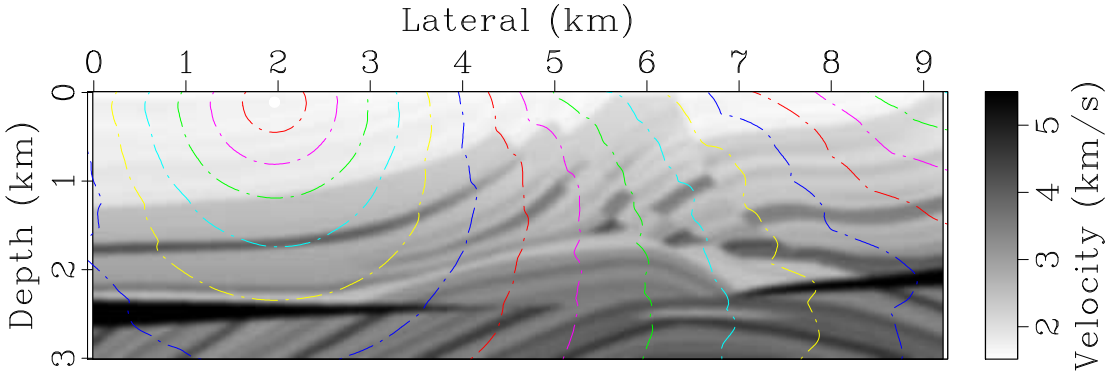
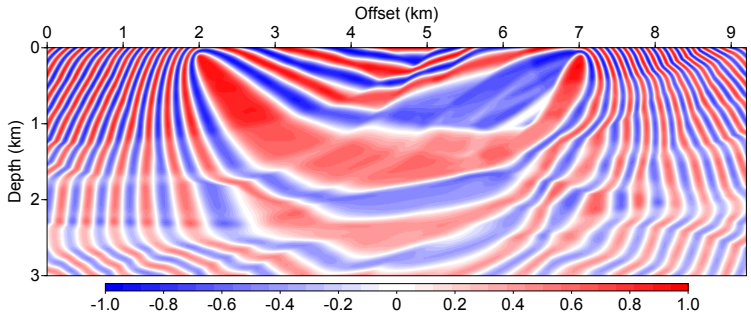


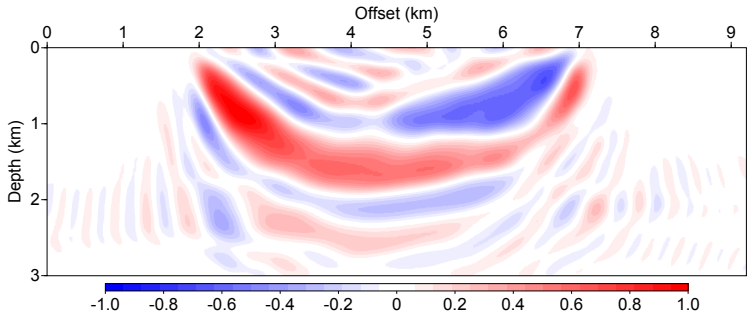
Figure 5: A slightly smoothed marmousi model using a rectangular smoother of 0.08 km in each direction. The traveltime solution from a source given by the white dot is given as contour plots.

. Since the extension is used to filter for low wavenumbers necessary for FWI convergence we have to consider rather long extensions in ζ which makes it very computationally intensive. The strong redundancy visible in the Figure 7 indicates that the filtering can be done more efficiently. The redundancy is even more visible when we do the Fourier transform to $(\mathbf{K},k_\zeta)$ domain. From the Figure 8 it seems that the Fourier transform is mostly zero outside some range along the k_ζ, moreover we prove in Appendix C that the image outside of the range should be exactly zero if we neglect the interpolation and finite extension artifacts, i.e.:

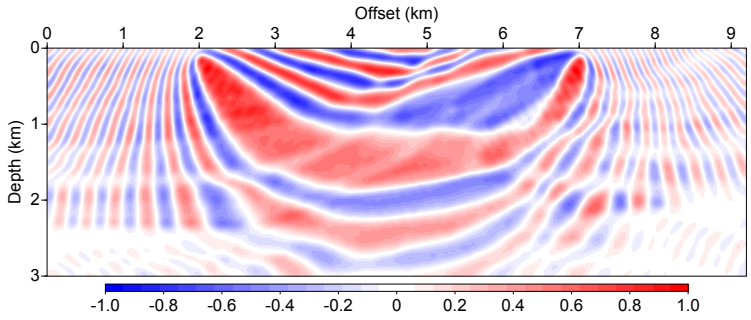
$$k_\zeta \in \left(\frac{2\omega}{v_{max}}, \frac{2\omega}{v_{min}}\right) \tag{11}$$



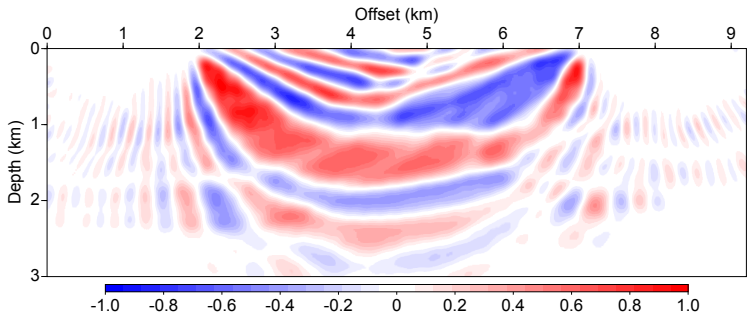
(a)



(b)



(c)



(d)

Figure 6: (a) Gradient (sensitivity kernel) for a single source and receiver, and a frequency of 7 Hz for a slightly smoothed Marmousi velocity model shown in Figure 5. The kernel in this example was computed using the ray-Born approximation, with the traveltime, as shown in Figure 5, calculated under the high-frequency asymptotic assumption. The gradients reflect the monochromatic nature of the wavefields used here, and we can clearly identify the low wavenumber wavepath parts of the gradient, as well as the scattering parts, given by the high wavenumbers. (b) Only deflection angles below 10 degrees are preserved with conventional filter. (c) Same as (b) but with the new filter. (d) Angles below 5 degrees with the new filter.

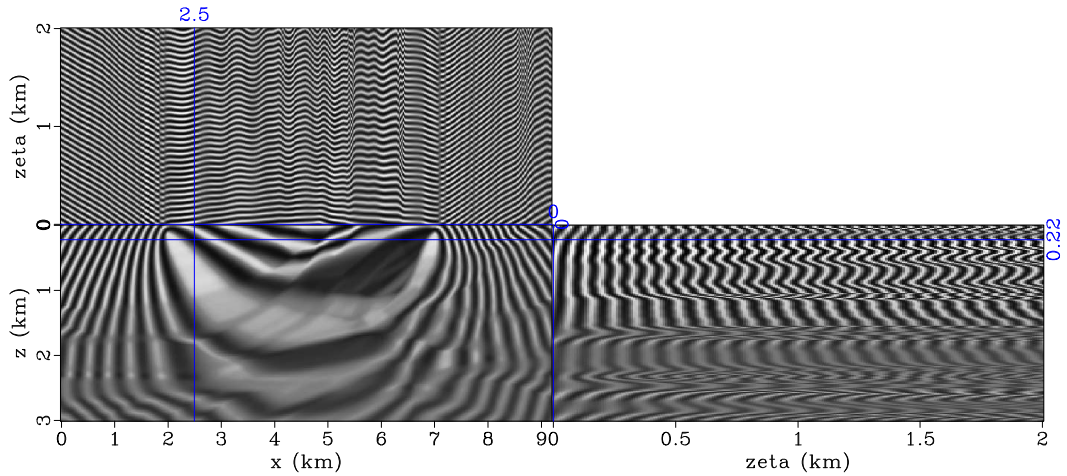


Figure 7: The ζ extended gradient based in equation 9 shows lots of redundancy in ζ which can be used to reduce the cost of the filter application.

Here v_{min} and v_{max} are the minimum and maximum velocities in the background model respectively. Thus the redundancy in the extended image can be reduced if it is computed for k_ζ within the range specified by equation 11.

As already mentioned within the conventional filtering process the image $I(\mathbf{K}, k_\zeta)$ (Figure 8) can be conditioned with equation 10 which results into a smoothed gradient version (Figure 6(b), 6(c), 6(d)) after applying inverse Fourier transforms and zero-lag imaging condition. For Figure 6(b) we selected the angle $\theta = 85^o$ which effectively restricts the frequencies used to 0.7 Hz, downscaling the frequency of 7 Hz 10 times.

For monochromatic gradients there is no real extension though. The image is simply multiplied by a coefficient varying with the background velocity. We show this on the example in Figure 6(a).

Figure 7 shows the extended image from the image given in Figure 6(a).

The efficient filtering algorithm

We propose to separate the original image into parts with different velocity ranges and process each of them as if the velocity was constant and equal to the average value in the respective range.

Thus $I(\mathbf{x})$ is composed of different summands based on the velocity at $\mathbf{x}$:

$$I(\mathbf{x}) = \sum I_i(\mathbf{x}), \tag{12}$$

where

$$I_i(\mathbf{x}) = I(\mathbf{x})U_i(\mathbf{x}), \tag{13}$$

$$U_i(\mathbf{x}) = \begin{cases} 1, \text{ if } v(\mathbf{x}) \in [v_i - \Delta v; v_i + \Delta v) \\ 0, \text{ if } v(\mathbf{x}) \notin [v_i - \Delta v; v_i + \Delta v). \end{cases} \tag{14}$$

The sequence of velocities is determined as follows:

$$v_{i+1} = v_i + 2\Delta v. \tag{15}$$

After the image is decomposed into areas with different velocities the scattering filter in each of them is applied as a stationary filter.

$$\tilde{I}_i(\mathbf{x}) = F^{-1} w(\mathbf{K}, v_i, \omega) F I_i(\mathbf{x}). \tag{16}$$

The filtered gradient is obtained by composing individual filtered images via equation 16:

$$\tilde{I}(\mathbf{x}) = \sum \tilde{I}_i(\mathbf{x}) \tag{17}$$

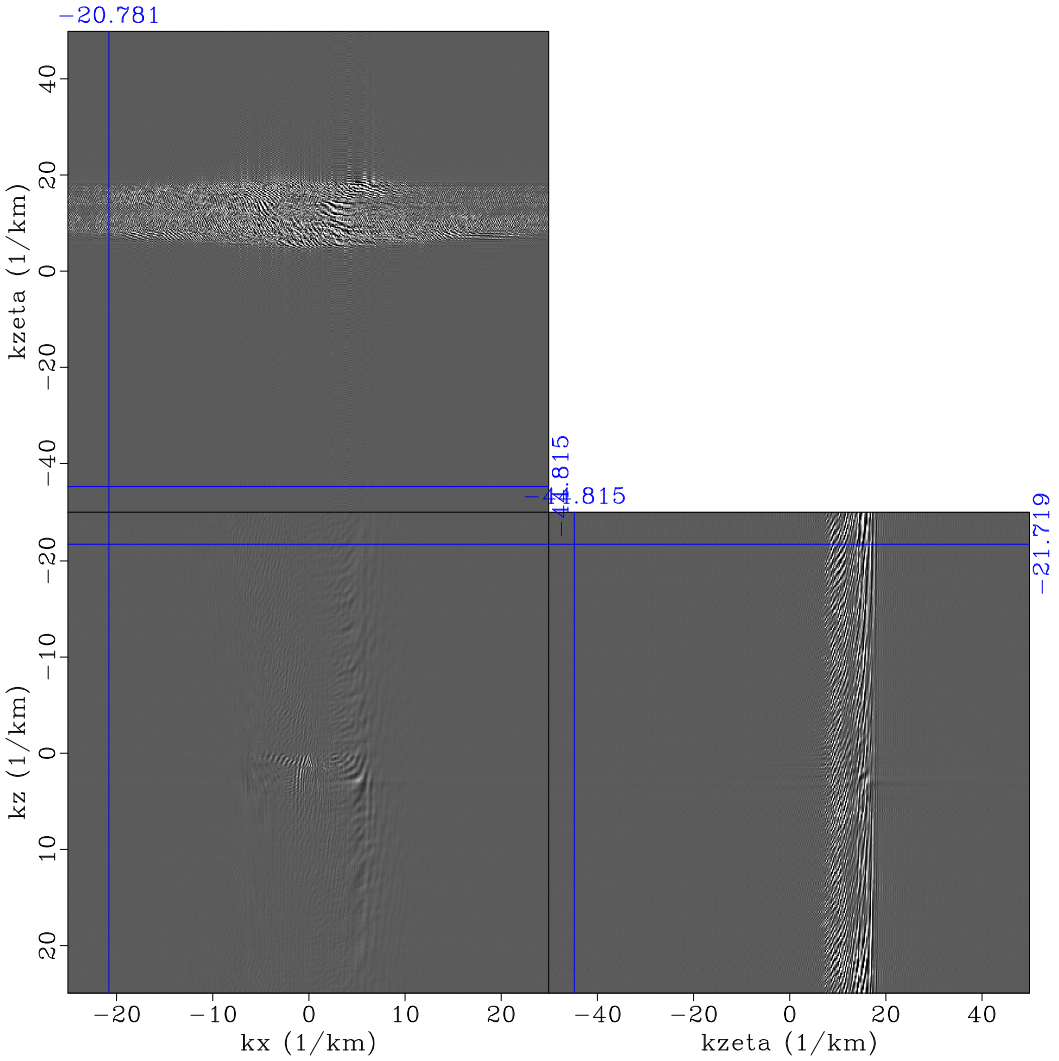
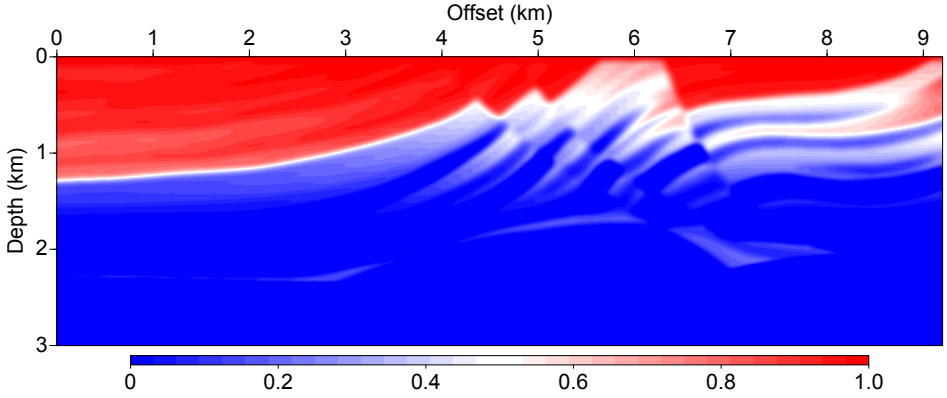


Figure 8: The 3-D spectrum of ζ extended gradient based in equation 9. The ζ extension shows that most of the energy is concentrated in the areas specified by the equation 11

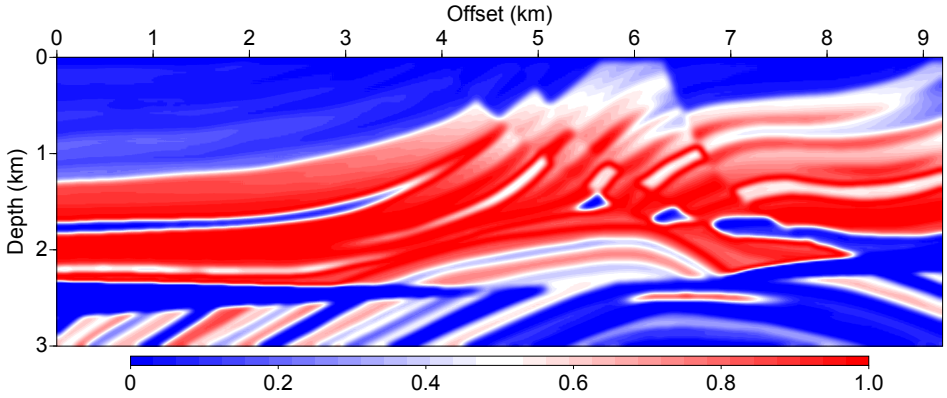
The total number of velocities v_i N, and Δv are adjusted so that the sequence of intervals covers the whole range of velocities in the background model. To obtain the functions U_i for a decomposition of unity the most straight forward is taking the characteristic functions for the intervals (equation 14). Yet this does not lead to good results at intermediate scattering angles due to the strong Gibbs phenomena. Thus we were forced to use the smooth version of a decomposition of unity expressed by equation 18.

$$U_i(\mathbf{x}) = \begin{cases} \cos^2\left(\frac{\pi(v(\mathbf{x})-v(v_i))}{4\Delta v}\right), \text{ if } v(\mathbf{x}) \in [v_i - 2\Delta v; v_i + 2\Delta v) \\ 0, \text{ if } v(\mathbf{x}) \notin [v_i - 2\Delta v; v_i + 2\Delta v). \end{cases} \tag{18}$$

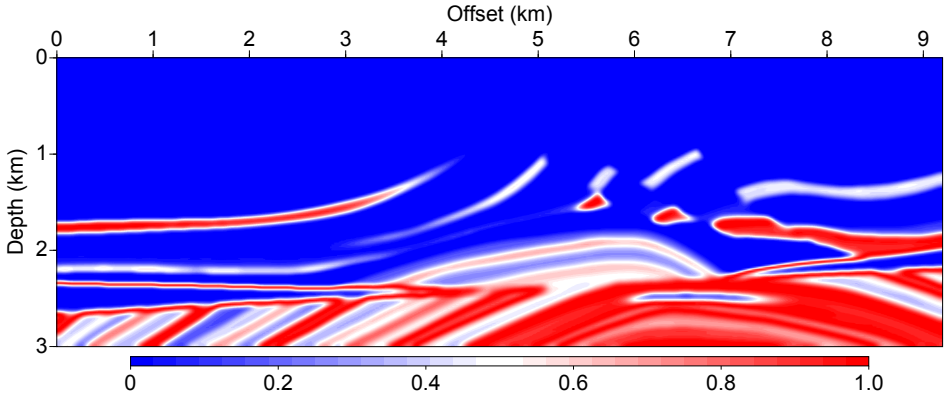
Our numerical examples show that a smooth decomposition leads to results very similar to those produced by the scattering angle filter with stretched time lag extension. We show how the model can be split on the example of the slightly smoothed Marmousi model to illustrate the general behavior of the filtering process in Figure 9(a). The total number of layers in this case is taken equal to 3. It shows how complicated domain boundaries of the velocity in given intervals can be.



(a)



(b)



(c)

Figure 9: The model decomposition functions U_i with smooth tapering. The smoothed Marmousi model is decomposed into 3 layers: (a) low velocity zones, (b) intermediate velocities, (c) high velocities.

DISCUSSION & CONCLUSIONS

The new filtering method described here takes the ray information and smears it over the model domain. This is what low frequencies tend to do over the first Fresnel zone, but low frequencies were not required here. The physical meaning of these filters is highlighted by isolating the energy that provides the spared scattering angles, and for a very narrow band around the transmission angle of 180 degrees, this corresponds to reducing the wavefields propagated from the source and receiver to local plane waves. As the narrow band widens slightly, these plane waves start to have a width controlled by the distance of the model point from the source and receiver respectively. In other words, some point source characteristics come into play. We actually have a phenomenon similar to what we encounter with beams for even a monochromatic wavefield.

There are a lot of weighting aspects that were ignored in the development of the update. Specifically, those extracted from the full Hessian. The diagonal part of the Hessian provides the appropriate geometrical spreading correction to the sensitivity kernel. Nevertheless, the contribution of the Hessian or its approximation will not alter the performance of the filtering. The filtering will still highlight the energy corresponding to the angles that we pass resulting in more localized energy along the ray path regions. What would differ is the distribution of energy within the filtered gradient.

In both strategies large offsets are desired to obtain smoother updates using the scattering angle filter, as well as to resolve anisotropy. These strategies are also meant for data dominated by near horizontal reflections.

ACKOWLEDGMENTS

We thank the WIT consortium, KAUST and G-RISC (German-Russian Interdisciplinary Science Center) for support. We thank the Seismic Wave Analysis Group and Applied Seismics Group Hamburg for many fruitful discussions. Vladimir Kazei is also grateful to Fons ten Kroode of Shell Global Solutions International BV for support of his PhD studies under CRDF grant RUG1-30020-ST-11 and Wim Mulder of TU Delft and Shell Global Solutions International BV for discussions.

REFERENCES

Alkhalifah, T. (2015a). Scattering-angle based filtering of the waveform inversion gradients. *Geophysical Journal International*, 200(1):363–373.

Alkhalifah, T. A. (2015b). Full model wavenumber inversion. *ASEG Extended Abstracts*, 2015(1):1–1.

Beydoun, W. B. and Mendes, M. (1989). Elastic ray-Born l2-migration/inversion. *Geophysical Journal International*, 97(1):151–160.

Biondi, B. and Almomin, A. (2013). Tomographic full-waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis. *The Leading Edge*, 32(9):1074–1080.

Bourgeois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P., and Versteeg, R. (1991). Marmousi, model and data. *The Marmousi Experience*, pages 5–16.

Bozdağ, E., Trampert, J., and Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2):845–870.

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

Devaney, A. J. (1984). Geophysical diffraction tomography. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(1):3–13.

Ewald, P. (1969). Introduction to the dynamical theory of X-ray diffraction. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 25(1):103–108.

Kazei, V. V., Troyan, V. N., Kashtan, B. M., and Mulder, W. A. (2013). On the role of reflections, refractions and diving waves in full-waveform inversion. *Geophysical Prospecting*, 61(6):1252–1263.

Khalil, A., Sun, J., Zhang, Y., and Poole, G. (2013). RTM noise attenuation and image enhancement using time-shift gathers. *SEG Technical Program Expanded Abstracts 2013*, 733:3789–3793.

Kosloff, D. D. and Baysal, E. (1982). Forward modeling by a Fourier method. *Geophysics*, 47(10):1402–1412.

Margrave, G. F. (1998). Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering. *Geophysics*, 63(1):244–259.

Mora, P. (1989). Inversion = migration + tomography. *Geophysics*, 54(12):1575–1586.

Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., and Dell'Aversana, P. (2004). Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. *Geophysical Journal International*, 159(3):1032–1056.

Rutherford, E. (1911). Lxxix. the scattering of α and β particles by matter and the structure of the atom. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(125):669–688.

Sava, P. C. and Fomel, S. (2003). Angle-domain common-image gathers by wavefield continuation methods. *Geophysics*, 68(3):1065–1074.

Sirgue, L. and Pratt, R. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.

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

Tessmer, E. (2011). Using the rapid expansion method for accurate time-stepping in modeling and reverse-time migration. *Geophysics*, 76(4):S177–S185.

van Leeuwen, T. and Herrmann, F. J. (2013). Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, 195(1):661–667.

Van Leeuwen, T. and Mulder, W. A. (2010). A correlation-based misfit criterion for wave-equation traveltime tomography. *Geophysical Journal International*, 182(3):1383–1394.

Virieux, J., Calandra, H., and Plessix, R.-É. (2011). A review of the spectral, pseudo-spectral, finite-difference and finite-element modelling techniques for geophysical imaging. *Geophysical Prospecting*, 59(5):794–813.

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

Wu, R.-S., Luo, J., and Wu, B. (2014). Seismic envelope inversion and modulation signal model. *Geophysics*, 79(3):WA13–WA24.

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

APPENDIX A

THE LOCAL SCATTERING ANGLE EQUATIONS

We start with the regular expression for the gradient of the regular quadratic misfit functional in the frequency domain for a single source-receiver couple:

$$\frac{\partial J}{\partial v}(\mathbf{x}) = \frac{\omega^2}{v(\mathbf{x})^3}\lambda(\mathbf{x},\omega)u(\mathbf{x},\omega). \tag{19}$$

Here u is the source wavefield and λ is the wavefield backpropagated from the receiver. In the frequency domain these two can easily be expressed with the use of Green's functions if we consider a single source-receiver experiment:

$$\frac{\partial J}{\partial v}(\mathbf{x}) = \frac{\omega^2}{v(\mathbf{x})^3}\delta u(\mathbf{x}_s,\mathbf{x}_r,\omega)G(\mathbf{x}_s,\mathbf{x},\omega)f(\omega)G(\mathbf{x}_r,\mathbf{x},\omega). \tag{20}$$

Equation (20) shows that the spatial distribution of the gradient is proportional to the product of two Green's functions, which we call the sensitivity Kernel. This product is fully defined by the background model, while the data residue is just a coefficient in the frequency domain.

$$\frac{\partial J}{\partial v}(\mathbf{x}) \propto G(\mathbf{x}_s,\mathbf{x},\omega)G(\mathbf{x}_r,\mathbf{x},\omega). \tag{21}$$

Equation (21) is apparently also suitable for RTM images analysis. The Green's functions can be decomposed into local monochromatic plane waves with local angle Fourier transforms:

$$G(\mathbf{x}_s,\mathbf{x}+\Delta\mathbf{x},\omega) \simeq \int_{|\mathbf{s}|=1} G(\mathbf{x}_s,\mathbf{x},\mathbf{s})\exp\left(i\frac{\omega}{v(\mathbf{x})}\mathbf{s}\cdot\Delta\mathbf{x}\right)\mathrm{d}\mathbf{s}. \tag{22}$$

Here the $\mathbf{s}$ is the direction of the local wavefield propagation in the plane wave decomposition of the source wavefield. $\mathrm{d}\mathbf{s}$ is an element of the 1-D unit circle in 2-D, and 2-D unit sphere in 3-D (we integrate in the angle domain which assumes that the non-homogeneous plane waves are neglected and the locally homogeneous media assumption is used). Analogously to the source wavefield we can decompose the receiver wavefield:

$$G(\mathbf{x}_r,\mathbf{x}+\Delta\mathbf{x},\omega) \simeq \int_{|\mathbf{r}|=1} G(\mathbf{x}_r,\mathbf{x},\mathbf{r})\exp\left(i\frac{\omega}{v(\mathbf{x})}\mathbf{r}\cdot\Delta\mathbf{x}\right)\mathrm{d}\mathbf{r}. \tag{23}$$

This gives the decomposition of the sensitivity kernel for the gradient of the misfit functional:

$$\frac{\partial J}{\partial v}(\mathbf{x}+\Delta\mathbf{x}) \propto \int\limits_{|\mathbf{s}|=1}\int\limits_{|\mathbf{r}|=1} \exp\left(i\frac{\omega}{v(\mathbf{x})}(\mathbf{s}+\mathbf{r})\cdot\Delta\mathbf{x}\right) \times G(\mathbf{x}_s,\mathbf{x},\mathbf{s})G(\mathbf{x}_r,\mathbf{x},\mathbf{r})\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{s}. \tag{24}$$

Now we can compare the equation (24) with the formula for the local Fourier transform of the gradient:

$$\frac{\partial J}{\partial v}(\mathbf{x}+\Delta\mathbf{x}) = \int \frac{\partial J}{\partial v}(\mathbf{x},\omega,\mathbf{K})\exp(i\mathbf{K}\Delta\mathbf{x})\mathrm{d}\mathbf{K}. \tag{25}$$

With an evident change of variables equation (24) can be rewritten in the fashion of equation (25). This leads to straightforward connection between the directions of illumination and the updated local wavenumber:

$$\mathbf{K} = \frac{\omega}{v(\mathbf{x})}(\mathbf{s}+\mathbf{r}) \tag{26}$$

We admit that this relation (26) should be used only for locally very slowly varying background velocities since the locally homogeneous assumption was used twice, first for the plane wave decomposition and second neglecting the variations of $1/v^3$ to take it out from the Fourier transform. Introducing the scattering angle into derivation we get the following relation for the absolute value of the updated wavenumber:

$$|\mathbf{K}| = \frac{2\omega}{v(\mathbf{x})}\cos(\theta/2). \tag{27}$$

APPENDIX B

DERIVATION OF THE GAUSSIAN PARAMETERS

The non-stationary Gaussian filtering is widely used in seismic data processing and sometimes even in full-waveform inversion (Ravaut et al., 2004). To get the benefits of the scattering angle based filtering we derive analytically a relation between the length of the Gaussian filter to the scattering angles kept. We start with the expression in the Fourier domain for the scattering angle filter which is nothing but the heaviside function tied to the local velocity.

$$\tilde{I}(\mathbf{x},\mathbf{K}) = H\left(\frac{2\omega}{v(x)}\cos(\theta/2) - |\mathbf{K}|\right) I(\mathbf{x},\mathbf{K}). \tag{28}$$

The Gaussian filter in the spectral domain at the same time acts like:

$$\tilde{I}(\mathbf{x},\mathbf{K}) = \exp\left(-\frac{\mathbf{K}^2}{2\sigma^2}\right) I(\mathbf{x},\mathbf{K}). \tag{29}$$

To fit to the original scattering angle filter we tune the Gaussian parameters to match the standard deviation of this filter. The standard deviation of the Gaussian filter is equal to σ. While the standard deviation of the distribution given by equation (28) can be evaluated in 2-D as follows:

$$\sigma_\theta^2 = \frac{\int_0^{2\pi}\int_0^{\kappa} l^3 \mathrm{d}l\mathrm{d}\phi}{\pi\kappa^2} = \frac{\kappa^2}{2}, \;\; \kappa \equiv \frac{2\omega}{v(\mathbf{x})}\cos(\theta/2). \tag{30}$$

Thus to mimic a scattering angle filter with maximum scattering angle θ one can take σ equal to σ_θ. To apply the filter one principally needs to apply the Fourier transform and Gaussian filtering for every $\mathbf{x}$ location. Fortunately the Fourier transform of the Gaussian function is also Gaussian ($\sigma_{\mathbf{x}} = \frac{1}{\sigma_{\mathbf{K}}}$) this allows us to get for each desired scattering angle based Gaussian filter an analytic formula in the spatial domain. This brings us to the formula:

$$\tilde{I}(\mathbf{x}) = \int \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{2\sigma_{\mathbf{x}}^2})I(\mathbf{x}',\mathbf{K}\right) \mathrm{d}\mathbf{x}' \tag{31}$$

APPENDIX C

ABOUT THE RELATION BETWEEN CONVENTIONAL EXTENSION AND THE NEWLY INTRODUCED VELOCITY EXTENSION

First, we rewrite equation 32:

$$I(\mathbf{x},\zeta) = I(\mathbf{x})e^{-2i\omega\frac{\zeta}{v(\mathbf{x})}}. \tag{32}$$

Applying the Fourier transform on ζ we obtain the following representation:

$$I(\mathbf{x},k_\zeta) = I(\mathbf{x})\delta(k_\zeta - \frac{2\omega}{v(\mathbf{x})}). \tag{33}$$

First of all the equation 33 suggest that for monochromatic data k_ζ is in fact a function of v. Second, equation 33 suggests that outside of the range $k_\zeta \in (\frac{2\omega}{v_{max}}, \frac{2\omega}{v_{min}})$ the spectrum of the extended image $I(\mathbf{K},k_\zeta)$ should be exactly zero. Yet this is not the case since the extension in ζ is discrete and finite, which causes smearing of the δ functions for all locations. One of the possible representations of the delta function which is used by the filter via extension with time lags is the discrete (fast) Fourier transform of the exponent in equation 33, while the other is some direct spatial approximation in the discrete k_ζ domain. The first way is the one proposed in (Khalil et al., 2013) and (Sava and Fomel, 2003).

The second is to estimate the delta functions to obtain directly $I(\mathbf{x},k_\zeta)$ to reduce the number of Fourier transforms leading to the $I(\mathbf{K},k_\zeta)$. Since we already know that in the image $I(\mathbf{x},k_\zeta)$, k_ζ represents the local velocity $v(\mathbf{x})$ (or more accurately the slowness variations) we can replace in the image the variable $I(\mathbf{x},v)$ with the same delta function representation.