Conditioning the full waveform inversion gradient to welcome anisotropy

Tariq Alkhalifah

CWP Research Report

April 21, 2014
Conditioning the full waveform inversion gradient to welcome anisotropy

Tariq Alkhalifah

ABSTRACT
Multi-parameter full waveform inversion (FWI) suffers from the complex nonlinearity in the objective function, compounded by the eventual tradeoff between the model parameters. A hierarchical approach based on frequency and arrival time data decimation to maneuver the complex nonlinearity associated with this problem usually falls short in anisotropic media. In place of data decimation, I use a model gradient filter approach to access the parts of the gradient more suitable to combat the potential nonlinearity and parameter trade off. The filter is based on representing the gradient in the time-lag normalized domain, in which small scattering angles of the gradient update are initially muted out. A model update hierarchical filtering strategy includes applying varying degree of filtering to the different anisotropic parameter updates. A feature not easily accessible to simple data decimation. Using both FWI followed by reflection based FWI (RFWI) is handled best by a combination of anisotropic parameters that starts with the horizontal velocity.

INTRODUCTION
The hierarchical approach of full waveform inversion (FWI) of starting with the low-frequency and early-arrival part of the data is slowly giving way to more emphases on filtering and conditioning the gradient (Tang et al., 2013; Almomin and Biondi, 2013; Albertin et al., 2013). This is especially true as we are trying to mix migration velocity analysis (MVA) and FWI (Xu et al., 2012; Ma et al., 2012; Almomin and Biondi, 2012; Fleury and Perrone, 2012; Wang et al., 2013). This is a natural progression that should remind us of our transformation in imaging analysis from flattening surface offsets to focusing on subsurface offsets. Despite the value of data decimation and selection, the real objective of this process is far more apparent in the model domain and specifically at the gradient level (Sirgue and Pratt, 2004). With MVA integrated into FWI, we rarely damp the later arrivals as they are needed for the MVA part, but the model domain provides the opportunity to select or shape the proper updates. For anisotropic media, model domain conditioning becomes even more important as we tend to have one source of data, but multi-parameters with their own gradient filtering requirements.

Multi-parameter inversion (Burridge et al., 1998; Plessix and Cao, 2011; Prieux...
et al., 2011) requires an appropriate choice of parameters to represent the model. The choice of parametrization will reduce the null space of the inversion only when we reduce the number of parameters we invert for. Finding a minimum set of parameters that can explain the data can lead to a better inversion. Alkhalifah and Plessix (2014) analytically analyzed the radial dependency (radiation pattern) of the anisotropic parameter perturbation in an acoustic transversely isotropic with vertical symmetry axis (VTI) media. They advocated using certain combinations of parameters for various FWI strategies, including those that start with a model obtained from MVA, and those dependent first on diving waves. Choosing the right inversion setup for resolving anisotropy can make the difference between interpretable high-resolution results and results that may not make a lot of sense (too smooth or too biased). By representing the VTI model using the NMO velocity, $\delta$, $\eta$ offers the proper perturbation radiation pattern for an inversion that includes reflections and diving waves. Since $\delta$ mildly influences the geometrical aspects of the recorded wavefield, it can serve as a secondary parameter to fit the amplitude to compensate for the shortcomings of the acoustic model in representing the true elastic Earth (a role that density plays in isotropic media). For an inversion with a hierarchical implementation in which diving waves are used first in the inversion, a VTI model represented by $v_n$, $\eta$, and $\epsilon$ offers a practical set necessary to reduce the tradeoff and provide reasonable resolution. In this case, $\epsilon$ plays the role of amplitude fitting as it mildly effects the kinematics in the recorded wavefield. The model's wavenumber resolution is generally similar to what we experience in isotropic media, but for the relation between the scattering angle and the model update, the wavenumber is different for the general VTI case.

In this paper, I extend the axis of the model update with a normalized time lag component capable of resolving the scattering angle, efficiently. This additional axis is used to properly filter the gradient to admit usable background updates for FWI from reflections and diving waves, as well as reflection full waveform inversion (RFWI). It utilizes the fact that a proper background update is not necessarily given by low frequencies or low wavenumbers, but more accurately by large scattering angles, where the update actually follows the rays.

**THE ANISOTROPIC RADIATION PATTERNS**

The radiation pattern, as opposed to the gradient wavenumber, provides the size of the influence, and specifically its directional dependence. For acoustic VTI media, Alkhalifah and Plessix (2014) derive such patterns for two sets of anisotropic parameter combinations that they deem to be the most practical. Those are $v_n$, $\eta$, and $\delta$ in the case in which we have the opportunity to resolve $v_n$ first, using for example migration velocity analysis (MVA), and $v_h$, $\eta$, and $\epsilon$ in the case in which we have the opportunity to invert for diving waves first. Both combinations utilize the fact that the kinematics of the wavefield as they are measured on the surface are dependent on $v_n$ or $v_h$ and $\eta$, with a disconnect to the weakly resolvable vertical velocity given by $\epsilon$ or $\delta$, used to fit the amplitude.
Using the asymptotic Green function (without multi-pathing),
\[ G(x, y, \omega) = A(x, y) \exp (ik_m \cdot x); \]
with \( k_m \) is the model update wavenumber given by \( k_m = \cos \frac{\theta}{2} m_0 \), where \( v_0 \) is the background velocity, and \( A \) is the geometrical amplitude. We can then write
\[ p_1(x_s, x_r, \omega) = -\omega^2 s(\omega) \int dx A(x_s, x, x_r, \omega) a_1(x) \cdot \mathbf{r}_1(x) \]
with \( s \) as the source function, and
\[ A(x_s, x, x_r, \omega) = \frac{G(x_s, x, \omega)G(x_r, x, \omega)}{v_0^2(x)\rho(x)}, \]
and
\[ \mathbf{r}_1 = \begin{pmatrix} r_{v_n} \\ r_\eta \\ r_\delta \\ r_\rho \end{pmatrix}; \quad a_1 = \begin{pmatrix} 2 \\ 2n_{sh}^2n_{rh}^2 \\ -(n_{sz}^2 + n_{rz}^2) \\ 1 + n_s \cdot n_r \end{pmatrix}. \]
The coefficients of \( a_1 \) define the radiation patterns of each parameter for the given parameterization, \( v_n, \eta, \delta, \rho \) (Aki and Richards, 1980).

The unit vectors \( \mathbf{n}_s \) and \( \mathbf{n}_r \) with the source incident angle, \( \theta_s \), and the reflector dip angle \( \alpha \), are given by
\[ \mathbf{n}_s = \begin{pmatrix} n_{sh} \\ n_{sz} \end{pmatrix} = \begin{pmatrix} \sin(\theta_s) \\ \cos(\theta_s) \end{pmatrix}; \quad \mathbf{n}_r = \begin{pmatrix} n_{rh} \\ n_{rz} \end{pmatrix} = \begin{pmatrix} -\sin(\theta_s + 2\alpha) \\ \cos(\theta_s + 2\alpha) \end{pmatrix}. \]
Since \( v_h = v_n\sqrt{1 + 2\eta} \) and \( 1 + 2\delta = \frac{1 + 2\varepsilon}{1 + 2\eta} \), we have the following relations between perturbations
\[ r_{v_n} = r_{v_h} - r_\eta; \quad r_\delta = r_\delta - r_\eta. \]
Thus, the radiation patterns of the parameterization \( (v_h, \eta, \varepsilon, \rho) \) are
\[ \mathbf{r}_2 = \begin{pmatrix} r_{v_h} \\ r_\eta \\ r_\varepsilon \\ r_\rho \end{pmatrix}; \quad a_2 = \begin{pmatrix} 2 \\ -n_{sh}^2n_{rh}^2 - n_{tz}^2 n_{sh}^2 \\ -(n_{sz}^2 + n_{sz}^2) \\ 1 + n_s \cdot n_r \end{pmatrix}. \]
(To obtain the radiation patterns, we use the relation \( n_{sh}^2 + n_{sz}^2 = 1 \).)

Considering the parameter combination of \( v_n, \eta, \) and \( \delta \), Figure 1(a) shows a vector plot with vector components given by the \( r_\eta \) (horizontal) and \( r_\varepsilon \) (vertical) \( (\eta \) and \( \delta \) radiation patterns, respectively) as a function of the scattering angle and reflector dip. The shaded regions represent the potential locations of reflections (bottom), diving waves (middle top), and RFWI (sides top) information. The size of the arrow (amplitude of the vector) is the size of the sensitivity with respect to perturbations in
\( \eta \) and \( \delta \), while the direction reveals the dependency distribution. For an arrow to point vertically implies that the sensitivity is due to a \( \delta \) perturbation. An arrow pointing horizontally implies an \( \eta \) dependency. For an arrow to point diagonally implies a tradeoff between the two parameters. On the other hand, the radiation pattern for the NMO velocity is invariant of direction, as given in equation 4, and thus has a tradeoff with \( \eta \) or \( \delta \) across the potential point illumination angles. As expected, the \( \eta \) resolution is strongest with diving waves and very large dips. However, in all cases, it inherits a tradeoff with \( v_n \). To invert for these model parameters, it would be wise to set \( \delta \), which does not affect the kinematics, and focus initially on \( v_n \) and \( \eta \).

On the other hand, for the parameter set \( v_n \), \( \eta \), and \( \epsilon \), Figure 1(b) shows that the dependency on both \( r_\eta \) (the vector’s horizontal component) and \( r_\epsilon \) (the vector’s vertical component) is weak in certain areas and specifically for diving waves, which makes the resolution of \( v_n \) possible, without setting any parameter constant. However, we also note that even if we resolve \( v_n \), we will need to set \( \epsilon \) initially (and hopefully close to its accurate value) constant to resolve \( \eta \) as the tradeoff between the two parameters is apparent everywhere. We can also invert for \( \epsilon \) first using reflections and then finally invert for \( \eta \) by accessing the larger angles or larger dips of reflection.

Instead of decimating the data to give the appropriate optimal gradient contribution for certain parameters, we can organize the model update by filtering the components we need at certain stages. In the next section, we consider an approach for doing so.

![Figure 1](image_url)

Figure 1: Vector plots depicting the sensitivity to \( \eta \) (horizontal component of the vector) and to either a) \( \delta \) or b) \( \epsilon \) (vertical component of the vector) for the combinations lead by \( v_n \) and \( v_n \), respectively.
FILTERING THE GRADIENT

As an alternative to the conventional hierarchal data selection approach used initially to isolate the long wavelength updates to the model (Sigue and Pratt, 2004), we can filter the gradient to provide such updates in spite of the data (Tang et al., 2013; Almomin and Biondi, 2013; Albertin et al., 2013). This recent development is based on the separation of scales in the update by applying the proper wavenumber filtering to the model update or matching the directional components of the source and receiver wavefields and, thus, decomposing the update to reflection and transmission components. In both cases, dipping reflections, in which the low wavenumber components are present in a direction not easily isolatable from the transmission, can cause problems. In fact, diving waves with limited source and receiver coverage may have high wavenumber components (at high frequencies) in certain directions, in spite of their value in updating the background, and specifically the path through which the wave traveled.

To improve the action of the gradient and specifically to allow it to focus on the appropriate long wavelength components, we can utilize an approach devised by (Khalil et al., 2013) to specifically filter out such components for a cleaner RTM image free of low-frequency artifacts, to do exactly the opposite and enhance these components. In line with Khalil et al. (2013), We introduce a slightly modified time lag (velocity scaled, ) to our conventional gradient (Alkhalifah, 2014),

\[
R_1(x, \zeta) = \sum_i u_{s_i}(x)u_{r_i}(x)e^{-\frac{4i\omega}{v}|x|},
\]

\[
R_2(x, \zeta) = \sum_i (u_{s_i}(x)\delta u_{r_i}(x) + u_{r_i}(x)\delta u_{s_i}(x))e^{-\frac{4i\omega}{v}|x|},
\]

where \(v\) is the velocity, and \(\zeta = \frac{2}{v}\theta(x)\). 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, \(\theta\), is then given by the following formula:

\[
\cos^2 \frac{\theta}{2} = \frac{|k_{m}|^2}{k_{\zeta}^2},
\]

where \(k\) is the wavenumber vector and \(k_{\zeta}\) is the wavenumber (Fourier transform) corresponding to \(\zeta\). A four-dimensional Fourier transform of \(\hat{R}(x, \zeta)\) (three-dimensional in 2D), will allow us to map \(\hat{R}(k, k_{\zeta})\) to its angle gather equivalence, \(\hat{R}(k, \theta)\) using equation 9. In our case, we use equation 9 to filter out the gradient energy corresponding to small \(\theta\) (reflections), starting probably with \(\theta < 170\) degrees, and just sum the rest over \(k_{\zeta}\) (the zero \(\zeta\) imaging condition). As a result, there is no need to map to angle gathers. Of course, we will have to inverse Fourier transform the image back to space to apply the gradient in space. The mute region to eliminate reflections in \(\hat{R}(k, k_{\zeta})\) is interestingly given by low conventional wavenumbers and high lag wavenumbers. Obviously, we also mute all energy laying in regions of \(\hat{R}(k, k_{\zeta})\), where \(\cos\frac{\theta}{2} > 1\).
Relating equation 9 to the model update wavenumber demonstrates that $k_\zeta = \frac{\omega}{v}$, which allows us to directly control the frequency content scaled by the velocity in the update. In other words, $k_\zeta$ controls the update wavenumber scale. Thus, the model update can be written:

$$k_m = k_\zeta \cos \frac{\theta}{2} n.$$  

This allows for a full control of the model update wavelength regardless of frequencies.

For minimum scattering angles of 100 and 160 degrees, the gray area in Figures 2(a) and 2(b), respectively, show the areas of the gradient represented in the space and $\zeta$ wavenumber domain that will be spared the muting. Below, we mute scattering angles lower than 179 degrees, which will provide a focus on the ray part of the gradient that admits actually almost zero wavenumber updates.

![Figure 2](image)

Figure 2: A plot showing in gray the regions in the model update (gradient) that be spared the muting to only allow a minimum scattering angle of a) 100 and b) 160 degrees. The highlighted regions correspond to reflections (bottom), to RFWI (upper corners), and diving waves (upper center).

**Gradient Wavenumber Distribution for Anisotropic Media**

Since a dip is typically treated as an unknown, and the Born linearized update is based on scattering theory (that is from a point in the model), the analysis here focuses on the dependency (radiation pattern) of the anisotropic parameters on the scattering angle and frequencies. We tend to use the dip to style or condition our gradient to emphasize the dip, but the constraints are not covered here.
Before we start filtering the anisotropic parameter gradients, let us gain some understanding of the wavenumber distribution of the sensitivity of the data to these parameters and learn about the tradeoff between the parameters in the domain we intend to filter in, specifically $k_\zeta$. From equation 9, we note that the scattering angle depends on the magnitude of $k_m$ and $k_\zeta$. Clearly, $k_m$ is given by the gradient function represented in the wavenumber domain. We therefore investigate the behavior of the anisotropic parameter sensitivity with respect to the dip, $\phi$, and $k_\zeta$. Figure 3(a) shows such dependency for the parameter combination given by $v_n$, $\eta$, and $\delta$. Like Figure 1(a), Figure 3(a) is a vector plot displaying the sensitivity magnitudes for $\eta$ (horizontal direction) and $\delta$ (vertical direction). The three highlighted regions correspond to our data, with the left block depicting the locations of conventional reflection (with reasonable maximum offset), the bottom right representing the location of diving waves, with their close to 180 degree scattering angles, and the top right corresponds to RFWI, where the effective dip of the kernel is more vertical for horizontal reflectors. For vertical reflectors, the region can extend to a zero dip as that part acts like a diving wave. However, for FWI, the region is reflection dependent. For the anisotropic combination led by $v_n$, the sensitivity is largest for $\delta$ for reflections in conventional FWI for small dips (vectors pointing vertically on the left side). The parameter $\eta$ takes over for reflections corresponding to large dips (vectors pointing horizontally on the left side) or large offsets. For diving waves, the data are sensitive to mainly $\eta$, and for RWI $\delta$ has the upper hand, unless the dips are large or the offset is extremely large, which induces smaller effective dip angles in the gradient, and $\eta$ starts to have more influence. For the $v_h$, $\eta$, and $\epsilon$ combination, the story changes (Figure 3(b)). For reflections in conventional FWI, the data are more sensitive to $\epsilon$ at small dips and $\eta$ starts to have influence at moderate dips, although both parameters have very mild influence at large dips. They also have very little influence on the diving wave part, which can be used solely to invert for $v_h$. For RFWI, there is a tradeoff between $\eta$ and $\epsilon$ in favor of $\epsilon$ for small offsets or small dips, and $\eta$ for larger offsets or dips.

A cut-off filter controlled by the scattering angle in which we mute a region of $k_\zeta$ below a certain value will exclude the classic FWI reflection contribution (left side of Figures 3(a) and 3(b)). Actually, in practical implementations, we should initially spare a small band at the right most part of $k_\zeta$. For the combination led by $v_n$, we seemingly would get $\eta$ from the diving waves, and $\delta$ from RFWI, both with tradeoff potential with $v_n$, which has an angular independent radiation pattern. For the combination led by $v_h$, we have little sensitivity to $\eta$ and $\epsilon$ for diving waves, but we have reasonable sensitivity to both parameters for RFWI.

GRADIENTS UNDER FILTERING IN ANISOTROPIC MEDIA

One of the virtues of the filtering in the model domain is that it can be applied to the model parameter updates separately, which is not easily done by decimating the
Figure 3: Vector plots synonymous to those shown in Figures ?? and 1(b) depicting the sensitivity to $\eta$ (horizontal component of the vector) and to either a) $\delta$ or b) $\epsilon$ (vertical component of the vector) for the combinations lead by $v_n$ and $v_h$, respectively. The highlighted regions correspond to FWI reflections (left), to RFWI (upper right), and FWI diving waves (lower left).

data over the frequency or offset, as these model parameters share the same data. For certain model parameters, the update can be exposed to excessive filtering, and we can loosen the filtering on other parameters to allow for higher-resolution contributions to the update. In the examples below, we consider the effect of filtering on classic FWI and RFWI gradients for perturbations in $\eta$ and $\delta$ or $\epsilon$. For the analysis of the scattering angle filter on $v_n$ and $v_h$ (both with angular invariant radiation patterns), I refer you to Alkhalifah (2014), which is devoted to isotropic media. We first consider the model parameterized by $v_n$, $\delta$ and $\eta$, a combination more suitable as discussed earlier for an inversion that relies first on RFWI. Thus, we show the RFWI gradients first.

**The $v_n$ combination**

Figure 4 shows the gradient in the $\eta$ perturbation for a monochromatic wavefield of 10Hz for the case of RFWI. Obviously, the components of the gradient corresponding to wavefields traveling nearly horizontal (i.e., a vertical reflector) have more energy, like the sides. On the other hand, the middle part of the gradient is dominated by having one of the wavefields (source or receiver or model perturbation) travel vertically, and thus the amplitude is small. Figure 5 shows the $\zeta$ extended version. The vertical section along $\zeta$ from the low energy middle part shows energy only at the perturbed model point. Figures 6(a) to 6(f) show the gradients after filtering
with various low-cut scattering angles. Despite the low amplitude contribution for the $\eta$ perturbation, the filtering managed to isolate such contributions along the wave path. As a result, we obtain the rabbit ears for certain cut-off angles, granted that the magnitude is small, compared for example to the $\delta$ perturbation as we will see later. This is inherent to the low sensitivity to perturbations in $\eta$ for such a setup. Larger offsets, providing rabbit ears with a wider opening, will admit more sensitivity to $\eta$, with similar behavior as a response to the filtering.

Figure 4: The RFWI model update response (sensitivity kernel) for the $\eta$ perturbation for a monochromatic wavefield of 10 Hz and a source located at 1 km and a receiver at 4 km, both at 0.1 km depth, and a model point at 2.5 km laterally and 2.5 km depth. The white dots correspond to the locations of the source, receiver, and model point perturbation.

Figure 5: The model update (sensitivity kernel) of Figure 4 with a $\zeta$ extension to allow for scattering angle identification.

In the case of $\delta$ perturbation, the gradient for RFWI for a single frequency of 10 Hz is shown in Figure 7. The response shows more energy in parts of the gradient corresponding to near vertical wavefield propagation, including those along the rabbit ears. Thus, the $\zeta$ extension shows energy in the vertical section from the middle crossing the perturbed model point (Figure 8). The application of scattering angle filtering shown in Figures 9(a) to 9(f) shows a similar response to that for the $\eta$ perturbations (Figures 6(a) to 6(f)), but with the opposite sign and more importantly a factor of magnitude higher. Thus, $\delta$ with this setup is better resolved than $\eta$. The opposite behavior happens with classical FWI.

After a proper, hopefully good enough for FWI, anisotropy model is attained from RFWI, we usually use that model as an initial model for classic FWI. The key requirement here is that the initial model can produce reflections at the available low frequencies that is a half cycle away from the observed data. However, since we are filtering the updates for FWI, we can relax such requirements as our focus...
Figure 6: The model update (sensitivity kernel) for RFWI of that in Figure 4 after applying the low-cut scattering angle filter muting angles below a) 179.4, b) 179, c) 178, d) 176, e) 170, and f) 160 degrees. The white dots correspond to the locations of the source, receiver, and model point perturbation.

Figure 7: The RFWI model update response (sensitivity kernel) for the δ perturbation for a monochromatic wavefield of 10 Hz and a source located at 1 km and a receiver at 4 km, both at 0.1 km depth, and a model point at 2.5 km laterally and 2.5 km depth. The white dots correspond to the locations of the source, receiver, and model point perturbation.
Figure 8: The model update (sensitivity kernel) of Figure 7 with a $\zeta$ extension to allow for scattering angle identification.

Figure 9: The model update (sensitivity kernel) for RFWI of that in Figure 7 after applying the low-cut scattering angle filter muting angles below a) 179.4, b) 179, c) 178, d) 176, e) 170, and f) 160 degrees. The white dots correspond to the locations of the source, receiver, and model point perturbation.
will still include the background model. The move from RFWI to FWI is mainly to address anisotropy. RFWI is used here to establish a good velocity model and FWI is hopefully ready to give us a good result. \( \delta \) as discussed previously will come from the higher-resolution amplitudes. Thus, to understand the effect of filtering on the updates of FWI, we consider again an isotropic constant velocity background medium. Figure 10 shows the classic FWI sensitivity kernel for the 10 Hz monochromatic wavefields. It is similar to the velocity one, but is has low energy for parts of the gradient corresponding to waves traveling vertically, like reflections, especially those with small scattering angles, as we saw in the previous chapter. Figure 11 shows the extended \( \zeta \) gradient. Figures 12(a) to 12(f) show the gradient response to scattering angle filtering. Unlike for RFWI, the \( \eta \) influence is large and the filter emphasizes that contribution along the ray. A strong low-cut scattering filtering will distribute the direct ray contribution over a large region centered between the source and receiver.

Figure 10: The model update response (sensitivity kernel) for the \( \eta \) perturbation for a monochromatic wavefield of 10 Hz and a source located at 1 km and a receiver at 4 km, both at 0.1 km depth, and a model point at 2.5 km laterally and 2.5 km depth. The white dots correspond to the locations of the source and receiver.

Figure 11: The model update (sensitivity kernel) of Figure 10 with a \( \zeta \) extension to allow for scattering angle identification.

On the other hand, the \( \delta \) gradient (Figure 13) has little energy between the source and receiver since that energy corresponds to horizontally traveling waves. The \( \zeta \) extension also emphasizes the lack of energy in the extended \( \zeta \) axis from the center. Since the \( \delta \) has almost no contribution especially along the line between the sources and receivers, the gradient filtering produces mixed results. Since there is no energy in the \( \delta \) perturbation, some of the cut-off angles produce low energy results void of direction. At very high-cut scattering angles like 179.5 and 179.7, we obtained extremely long wavelength energy, but in the right direction update. Otherwise, the response has a high wavenumber, which is possibly not helpful in avoiding the local
Figure 12: The model update (sensitivity kernel) for the classic inversion of that in Figure 10 after applying the low-cut scattering angle filter muting angles below a) 179.4, b) 179, c) 178, d) 176, e) 170, and f) 160 degrees. The white dots correspond to the locations of the source and receiver.
minima. Thus, for $\delta$, we can keep the high-cut filter until a proper background is calculated for $v_n$ and $\eta$.

Figure 13: The model update response (sensitivity kernel) for the $\delta$ perturbation for a monochromatic wavefield of 10 Hz and a source located at 1 km and a receiver at 4 km, both at 0.1 km depth, and a model point at 2.5 km laterally and 2.5 km depth. The white dots correspond to the locations of the source and receiver.

Figure 14: The model update (sensitivity kernel) of Figure 13 with a $\zeta$ extension to allow for scattering angle identification.

In all cases, since the NMO velocity (the third parameter in this combination) has a stationary radiation pattern with an angle, there will be a tradeoff with both parameters (Alkhalifah and Plessix, 2014). This means, in this setup, that if we set $\delta$ and $\eta$ to some stationary value, our inversion for $v_n$ will be dependent on the accuracy of that value. In the parametrization, led by the horizontal velocity, we have a region in the radiation pattern that is exclusively reserved for $v_h$.

**The $v_h$ combination**

In this combination as suggested by Alkhalifah and Plessix (2014), the model is described by $v_h$, $\eta$, and $\epsilon$. Since, as Figure 3(b) shows, the horizontal velocity has practically no tradeoff with the other two parameters for diving waves, we start the inversion and our analysis with the classic FWI.

Figure 16 shows the classic FWI sensitivity kernel for the 10 Hz monochromatic wavefields for this combination. Unlike in the previous model representation, the data is less sensitive now to $\eta$, as $v_h$ mainly describes the data for horizontal propagation. Figure 17 shows the extended $\zeta$ gradient. Figures 18(a) to 18(f) show the gradient response to scattering angle filtering. Unlike for the $v_n$ combination, the $\eta$ influence is small and the filtering indicates the lack of contribution along the ray. In fact,
Figure 15: The model update (sensitivity kernel) for the classic inversion of that in Figure 13 after applying the low-cut scattering angle filter muting angles below a) 179.4, b) 179, c) 178, d) 176, e) 170, and f) 160 degrees. The white dots correspond to the locations of the source and receiver.
since $\epsilon$ has a similar response as $\delta$, shown in the previous subsection, $\eta$ and $\epsilon$ mildly effect the data for direct waves. Such data can be used to invert exclusively for $v_h$ and we can keep both $\eta$ and $\epsilon$ at with a high low-cut scattering angle filter, which admits only smooth updates. On the other hand, the $\epsilon$ gradient is exactly similar to

Figure 16: The model update response (sensitivity kernel) for the $\eta$ perturbation for a monochromatic wavefield of 10 Hz and a source located at 1 km and a receiver at 4 km, both at depth 0.1 km, and a model point at 2.5 km laterally and 2.5 km in depth. The white dots correspond to the locations of the source and receiver.

Figure 17: The model update (sensitivity kernel) of Figure 16 with a $\zeta$ extension to allow for scattering angle identification.

the $\delta$ one from the previous section as they have the same radiation pattern for the different combinations.

As soon as a good $v_h$ model is extracted from direct and diving waves, we can start using RFWI to get a good background model for $\eta$. Figure 19 shows the gradient in $\eta$ perturbation for a monochromatic wavefield of 10 Hz for the case of RFWI. Now, unlike in the $v_n$ combination, we have energy in the center. Figure 20 shows the extended $\zeta$ version. Figures 21(a) to 21(f) show the gradient after filtering with various low-cut scattering angles. The filtering managed to lower the wavenumber of the gradient and localize it to the potential wave paths. As a result, we obtain the rabbit ears for certain cut-off angles. As soon as a viable smooth $\epsilon\eta$ is obtained, we return back to FWI to resolve the high wavenumber components of the velocity model by relaxing the scattering angle filter. For the $v_h$ parametrization, $\eta$ has more influence at these angles than it had in the $v_n$ parametrization.

**DISCUSSION**

The new filtering method described here takes the ray information and smears it over the model domain. This is what low frequency tends to do over the Fresnel zone, but no low frequency was required here. The physical meaning of these filters is
Figure 18: The model update (sensitivity kernel) for the classic inversion of that in Figure 16 after applying the low-cut scattering angle filter muting angles below a) 179.4, b) 179, c) 178, d) 176, e) 170, and f) 160 degrees. The white dots correspond to the locations of the source and receiver.
Figure 19: The RFWI model update response (sensitivity kernel) for the $\eta$ perturbation for a monochromatic wavefield of 10 Hz and a source located at 1 km and a receiver at 4 km, both at 0.1 km depth, and a model point at 2.5 km laterally and 2.5 km depth. This $\eta$ perturbation is based on the model parametrization given by $v_h$, $\eta$, and $\epsilon$. The white dots correspond to the locations of the source, receiver, and model point perturbation.

Figure 20: The model update (sensitivity kernel) of Figure 19 with a $\zeta$ extension to allow for scattering angle identification.
Figure 21: The model update (sensitivity kernel) for RFWI of that in Figure 19 after applying the low-cut scattering angle filter muting angles below a) 179.4, b) 179, c) 178, d) 176, e) 170, and f) 160 degrees. The white dots correspond to the locations of the source, receiver, and model point perturbation.
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 wavefield from the source and receiver to plane waves. As the narrow band widens slightly, the plane waves start to have a width controlled by the distance of the model point from the source and receiver. 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.

We can summarize the approaches described above into two main strategies to address anisotropy in FWI. The first one is for the case where we have large offsets and credible diving waves. As discussed earlier, in this case, I recommend using the combination $v_n$, $\eta$, and $\epsilon$. Figure 22(a) shows the strategy for filtering, for this case, using both RFWI and FWI. However, if large offsets, or credible diving waves are not available we have to resort to the second strategy outlined in Figure 22(b), which tends to have more trade off between the parameters. In this case, we invert for $v_n$, $\eta$, and $\delta$. In both cases, we are expect to end up with a high resolution velocity model, a smooth $\eta$ (an inherent limitation of our surface acquisition (Djebbi et al., 2014)), and a high resolution erroneous $\epsilon$ or $\delta$. In every step, the scattering angle filtering starts with high angles down to low angles. 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.

**CONCLUSIONS**

Despite the many features that filtering the gradient brings to the table in the isotropic case, the benefits for anisotropic media are even larger. It allows us to apply different filtering for different parameters. We can apply a specific filter for FWI and another for RFWI, with specific features corresponding to the parameters of perturbation. For an acoustic anisotropic medium, in which we invert for three anisotropic parameters, using FWI and RFWI allows for six different filtering strategies. These strategies depend on the data content including the amount of diving waves present in the data. Such strategies can be formulated from the analysis provided here.
Figure 22: Work flows for applying scattering angle filters to FWI and RFWI a) for the case of usable diving waves in the data, and thus, inverting for a combination given by \( v_h, \eta, \) and \( \varepsilon, \) and b) for the case where horizontal reflectors are dominant in which diving waves are poorly recorded, and thus, inverting for a combination given by \( v_n, \eta, \) and \( \delta. \)

ACKNOWLEDGMENTS

I thank Rene Plessix and Ramzi Djebbi for their fruitful comments and suggestions. I also thank Ilya Tsvankin and Dave Hale for many useful discussions.

REFERENCES


———, 2013, Tomographic full waveform inversion (tffi) by successive linearizations


