Extended full waveform inversion with matching filter

Yuanyuan Li and Tariq Alkhalifah

Physical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; yuanyuan.li@kaust.edu.sa (Corresponding author); tariq.alkhalifah@kaust.edu.sa.

Keywords: Full-waveform inversion, Seismic velocities, Acoustic.

ACKNOWLEDGEMENTS

We would like to thank KAUST for its support and the members of Seismic Wave Analysis Group (SWAG) for their helpful discussions. The Shaheen supercomputing Laboratory in KAUST provides the computational support. The real data shown in this study are provided by courtesy of CGG.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1365-2478.13121.

This article is protected by copyright. All rights reserved.
Conflict of Interest Statement

The authors have no conflicts of interest to declare.

All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

ABSTRACT

Full waveform inversion has shown its huge potentials in recovering a high-resolution subsurface model. However, conventional full waveform inversion usually suffers from cycle skipping, resulting in an inaccurate local minimum model. Extended waveform inversion provides an effective way to mitigate cycle skipping. A matching filter between the predicted and observed data can provide an additional degree of freedom to improve the data fitting and avoid the cycle skipping. We extend the search space to treat the matching filter as an independent variable that we use to bring the compared data within a half cycle to obtain accurate direction of velocity updates. We formulate the objective function using the penalty method by linearly combining a data-misfit term and a penalty term. The objective function with a reasonable penalty parameter has a larger region of convergence compared to conventional full waveform inversion. We search the optimal solution over the extended model by updating the matching filter and the velocity in a nested way. The normalization of the data can bring us an equivalent normalization to the filter, and a more effective convergence. In the synthetic Marmousi model, the proposed inversion method recovers the
velocity model stably and accurately starting from a linearly increasing model in the case of lack of low frequencies below 3 Hz, in which conventional full waveform inversion suffers from cycle skipping. We also use a marine field data to demonstrate the effectiveness and practicality of the proposed method.

**Keywords:** Full-waveform inversion, Seismic velocities, Acoustic.

**INTRODUCTION**

Full waveform inversion (FWI) provides an effective and quantitative way to estimate the subsurface properties by minimizing the misfit between the predicted and observed data (Tarantola 1984). It is known that this methodology is proposed and developed in exploration geophysics originally (Tarantola 1984) and has been expanded to other image or detection fields, such as ground-penetrating radar (Busch et al. 2014) and medical imaging (Sandhu et al. 2015). Conventional FWI compare the data point-by-point using least-squares misfit function. However, when the initial model can’t predict the data within a half period, conventional FWI tends to surfer from cycle skipping and be trapped into a meaningless local minimum (Virieux and Operto 2009). With the development of high-performance computing and wide-azimuth acquisition, FWI has been widely studied in the last decade. A variety of approaches have been proposed to alleviate the cycle skipping of FWI (Ma and Hale 2013; Alkhalifah 2015; Warner and Guasch 2016; Métivier et al. 2016; Wu and Alkhalifah 2018; Li et al. 2019; Yao et al. 2019; Song et al. 2020; Li and Alkhalifah 2020;). It is intuitive to select part of the data free from cycle skipping as the matching objective. The most widely used strategy is a hierarchical inversion of frequencies, offsets and wave types. Among them,
the multiscale strategy aims at improving the resolution gradually by introducing progressively the high-frequency components (Bunks et al. 1995; Sirgue and Pratt 2004; Li et al. 2018). Besides, the refraction or diving waves can serve as the input of FWI considering the data misfits of them show better convexity and contains fewer local minima than that of the full data set (Shen 2014; Choi and Alkhalifah 2018; Liu et al. 2018). However, manual intervention for quality control is always required during the inversion process.

It is well known that conventional FWI based on L2-norm misfit function will fall into a local minimum when the starting model is not located in the convergence region of global minimum. Therefore, considerable efforts have been devoted to build a promising objective function with better convexity. The objective functions using an envelope (Wu et al. 2014; Oh and Alkhalifah 2018), auxiliary bump functional (Bharadwaj et al. 2016) and nonlinearly smoothed wavefield (Li et al. 2018) show promising inversion results even in the case of a lack of low frequencies because the high-frequency oscillations in the band-limited seismic data are suppressed and in favor of artificially produced low-frequency information.

Extended waveform inversion is a popular way to overcome the limitations of FWI by extending the search domain, which enlarges the model space with nonphysical degrees of freedom (Symes 2008). The introduction of an additional nonphysical dimension to the model makes it possible that the simulated data match the observed data well, even with kinematically inaccurate velocity models, and thus cycle skipping can be overcome (Symes 2008). The subsurface offset (Lameloise and Chauris 2016; Fu and Symes 2017; Dafni and Symes 2019), time lag (Yang and Sava 2013; Biondi and Almomin 2014), reconstructed wavefields (van Leeuwen and Herrmann 2013; Alkhalifah and Song 2019), or a matching filter (Warner and Guasch 2016; Sun and Alkhalifah 2019a)
can all serve as extensions providing additional degrees of freedom. They can improve the data fitting and extend the convergence domain of the objective function. In general, the matching filter for each data trace is computed by solving a 1-D linear inverse problem. Thus, the matching filter based inversion is more practical in terms of computational cost.

Deconvolution plays an important role in obtaining a high-quality subsurface image which is usually distorted by band-limited source wavelet (Mendel 2013). This technology has been utilized in FWI to build an objective function with better performance. Deconvolution-based objective function (Luo and Sava 2011) is optimized iteratively during inversion to drive the energy in the matching filter computed by deconvolution to concentrate at zero-time lag. This objective function shows less sensitivity to bandlimited or nonimpulsive sources compared to the crosscorrelation-based objective function (van Leeuwen and Mulder 2010). Warner and Guasch (2016) employ normalization on the matching filter and penalized the normalized matching filter evaluated by deconvolving the observed and predicted data to find the velocity model that renders the matching filter to be a Dirac delta function, which is referred to as adaptive waveform inversion (AWI). Sun and Alkhalifah (2019a) utilized the Wasserstein distance in the optimal-transport theory to compute the distance between the matching filter in the Radon domain and an approximate zero-lag delta function. In their approach, the matching filter should be converted to a distribution by a normalization term to adhere to the requirement of Wasserstein distance. Huang et al. (2017) used an extended nonphysical source for each trace through regularized deconvolution, and then forced the nonphysical source toward a physical one. For complex wavefields with multipath arrivals, the matching filter solved by a direct deconvolution will involve a slowly decaying signal with energy at time lags unrelated to traveltime error. Huang et al. (2017) applied Tikhonov regularization to
constrain the deconvolution in the frequency domain. They used a reduced objective function, which is uniquely determined by the velocity model, to remove the nonphysical part from the model progressively. As Fu and Symes (2017) reported, the objective function of extended waveform inversion is supposed to aim to minimize the extent of the model extension in the inversion process, while simultaneously minimizing a measure of data misfit. In this paper, we combine a data-misfit term and a penalty term about the matching filter to formulate a penalty function over the extended model: the matching filter and the velocity model. The penalty term is a natural regularization for the ill-posed inverse problem (deconvolution) by suppressing the large-lag signal. The matching filter and velocity model are two independent inversion parameters, and they are updated in a nested way to optimize the penalty function. Because the inversion involves both the matching filter and the model parameter, the search space is enlarged, and thus, the local-minimum problem can be mitigated. To avoid simply decreasing the coefficients of the filter to minimize the function, we normalize the observed and modeled data to obtain an equivalent normalization of the filter and maintain a linear optimization problem for the matching filter.

In the following, we start with a review of conventional FWI and the application of the matching filter in FWI and then present a novel formulation for matching filter-based extended FWI using the penalty method. In the example section, we use a modified Marmousi model to illustrate the performance of the proposed objective function and analyse the impacts of the penalty parameter and normalization on the inversion performance. Then, we test the proposed inversion approach on the synthetic Marmousi model and a marine field data. Finally, we discuss the limitations and possible improvements of the proposed method.
THEORY

Review of conventional FWI

Full waveform inversion aims to estimate the subsurface model by minimizing the misfit between the modeled data \( u(m) \) and observed data \( d \). The least-square objective function is defined as:

\[
\min_{m} J(m) = \frac{1}{2} \| u(m) - d \|^2.
\]  

(1)

We simulate the propagation of seismic waves by solving an acoustic wave equation using the finite-difference method and extract the wavefield at the receivers as the modeled data \( u(m) \).

We update the subsurface model \( m \) iteratively to minimize the objective function using the following equation:

\[
m_{n+1} = m_n - \alpha_n H^{-1} \nabla_m J(m),
\]

(2)

where, the subscript \( n \) is the iteration number, \( \alpha_n \) is the step length, \( H^{-1} \) refers to an approximate inverse Hessian matrix. The gradient \( \nabla_m J(m) \) of the objective function with respect to the subsurface model is given by

\[
\nabla_m J(m) = \frac{\partial J(m)}{\partial m} = \left( \frac{\partial u(m)}{\partial m} \right)^T (u(m) - d).
\]

(3)
Using the adjoint-state method (Plessix 2006), the gradient can be computed by cross-correlating the forward-propagated wavefield with the back-propagated wavefield from the adjoint source \((u(m) - d)\) at the receivers.

**FWI using matching filter**

Conventional FWI measures the data misfit in a point-by-point manner with the least-squares objective function and tends to be a highly nonlinear inverse problem. The matching filter, which is computed by a deconvolution of modeled and observed data, provides an alternative way to measure the data misfit. If the model is exact, the matching filter is supposed to reduce to a zero-lag Dirac delta function in the time domain and unity in the frequency domain. In this case, the energy in the matching filter is focused on zero time lag. Therefore, we define the objective function as the squared L2-norm of weighted matching filter \(w\) (Luo and Sava 2011):

\[
\min_{m} J(m) = \frac{1}{2} \|Tw\|^2 ,
\]

where, the weighting factor in the diagonal matrix \(T\) gradually increases with temporal lag to penalize the energy in the matching filter \(w\) away from zero lag. The matching filter \(w\) serves to match the modeled \(u\) to the observed data \(d\) in a least squares sense, and it satisfies the convolutional (*) formula:

\[
u(m) * w = d .
\]
When the model \( \mathbf{m} \) is exact, that is \( \mathbf{u}(\mathbf{m}) \) matches \( \mathbf{d} \) well, the matching filter \( \mathbf{w} \) often reduces to an approximate zero-lag delta function. The diagonal weighting matrix \( \mathbf{T} \) is used to force the filter energy towards zero lag. Correspondingly, the inversion can move towards the global optimal model. However, the direct deconvolution for \( \mathbf{w} \) often suffers from ill-posedness or instability, and the matching filter is contaminated with decaying oscillation signal, which fails to indicate the similarity between \( \mathbf{u}(\mathbf{m}) \) and \( \mathbf{d} \). A popular remedy to this ill-posed inverse problem is the Tikhonov regularization (Engl et al. 1996). The matching filter can be solved either in the frequency domain (Luo and Sava 2011; Huang et al. 2017; Sun and Alkhalifah 2019a), or in the time domain as below:

\[
\mathbf{w} = (\mathbf{U}^T \mathbf{U} + \varepsilon \mathbf{I})^{-1} \mathbf{U}^T \mathbf{d},
\]  

(6)

where, \( \mathbf{U} \) is a Toeplitz matrix representing convolution by \( \mathbf{u}(\mathbf{m}) \), \( \varepsilon \) is a regularization factor to stabilize the deconvolution.

**Matching filter-based extended FWI**

The abovementioned constrained optimization problem (equations 4 and 5) can be recast as an unconstrained problem using the penalty method. In this way, Huang et al. (2017) introduced an objective function for source-based extended waveform inversion in the frequency domain without normalization.

The normalization has been widely used in the crosscorrelation-based objective function to alleviate the dependence on the amplitude in matching the data (Choi and Alkhalifah 2013). Warner and Guasch (2016) normalized the filter to avoid the amplitude
trade-off between the filter and data, underlying the AWI algorithm. Sun and Alkhalifah (2019b) measure the focusing of the matching filter using Wasserstein W2 distance, which is used to calculate the optimal transport between two probability distribution functions. Therefore, the normalization is naturally introduced to transform the matching filter to a probability distribution function. The normalization term shows vital potential to improve the convexity of the objective function and accelerate its convergence (Warner and Guasch 2016; Sun and Alkhalifah 2019b). Thus, we also normalize the two data sets in our objective function (Li and Alkhalifah 2019).

Similar to the classic extended domain approach (Symes 2008), we linearly combine a data-misfit term and a penalty term to build the objective function:

$$\min_{\hat{m}, \hat{w}} J_\lambda(\hat{m}, \hat{w}) = \frac{1}{2} \left\| \hat{u}(m) * w - \hat{d} \right\|^2 + \frac{\lambda}{2} \left\| Tw \right\|^2,$$

where,

$$\hat{u}(m) = \frac{u(m)}{\|u(m)\|},$$

$$\hat{d} = \frac{d}{\|d\|}.$$

The diagonal elements of $T$ are set as the absolute value of time lag for simplicity. The penalty parameter $\lambda$ controls the extent of penalty for model extension or any departure from a zero-lag Dirac delta function. When the penalty parameter $\lambda \to 0$, no penalty is imposed on the model extension, i.e., the matching filter, that is to say, the model extension favors fully the data fitting. However, the sub-problem corresponding to $w$ is a direct deconvolution without regularization, which tends to be unstable for multiple arrivals, and
thus, the resolved $w$ fails to indicate the data similarity. When the penalty parameter $\lambda \to \infty$, we can’t benefit from the extended search space any more.

Considering $\lambda$ plays an important role in helping us avoid cycle skipping, we will show the performance of the objective function with different $\lambda$ in the next section. We pick the penalty parameter $\lambda$ using trial-and-error to balance the data fitting and the regularization on the deconvolution. In the examples, we choose a value for $\lambda$ that can reduce the data error to approximately 50%.

The penalty function (equation 7) entails optimizing over the matching filter $w$ and the subsurface model $m$, and thus, exploits a larger search space. We break down the joint optimization problem over the extended space into a sequence of two sub-problems corresponding to $w$ and $m$, respectively. The matching filter and the subsurface model are updated in a nested way. Specially, we first invert for the filter $w$ with a fixed model $m$. Given the inverted filter $w$, we can update the model $m$ once to further reduce the data misfit. The optimization problem (equation 7) is recast as two sub-problems regarding $w$ and $m$ that are solved in sequence:

$$w^{k+1} = \arg \min_w \frac{1}{2} \|\hat{u}(m^k) * w - \hat{d}\|^2 + \frac{\lambda}{2} \|Tw\|^2,$$  \hspace{1cm} (10)

$$m^{k+1} = \arg \min_m \frac{1}{2} \|\hat{u}(m) * w^{k+1} - \hat{d}\|^2,$$  \hspace{1cm} (11)

The optimization problem defined by equation 10 has a closed-form solution:

$$w = (U^T U + \lambda T^T T)^{-1} U^T d,$$  \hspace{1cm} (12)
We presume that the matrix $U^T U + \lambda T T^T$ is positive definite, while it’s hard to adhere to this condition, especially in the case that $\lambda$ is too large or too small. The linear least-square problem (equation 10) has a single global minimum and it can be solved by an iterative method. The gradient for the matching filter $w$ can be derived by taking the derivative of the objective function with respect to $w$:

$$g_w = \hat{u}(m^k) \otimes \left( \hat{u}(m^k) \ast w - \hat{d} \right) + \lambda T^2 w,$$

where, $\otimes$ represents the cross-correlation operator.

Substitute $w = w - \alpha_w g_w$ into equation 10, we determine the optimal step length for $w$ ($\alpha_w$) as a value that makes the derivative of the objective function with respect to $\alpha_w$ equal zero.

The updating step length for $w$ is, thus, given by:

$$\alpha_w = \frac{\left( \hat{u}(m^k) \ast w - \hat{d} \right) \left( \hat{u}(m^k) \ast g_w \right) + \lambda T w \cdot T g_w}{\left( \hat{u}(m^k) \ast g_w \right)^2 + \lambda \left( T g_w \right)^2}.$$

Given the gradient and the step length, we can update the matching filter iteratively using a conjugate gradient method. With the inverted matching filter, we then update the model $m$ to minimize the objective function shown in equation 11. The gradient of this function with respect to $m$ is computed by the following equation 15, and the detailed derivation is shown in Appendix.

$$g_m = \frac{\partial J}{\partial m} = \frac{\hat{u}(m^{k+1}) \otimes \left( \hat{u} \ast w^{k+1} - \hat{d} \right)}{\left\| w^{k+1} \otimes \left( \hat{u} \ast w^{k+1} - \hat{d} \right) - \hat{u} \left[ w^{k+1} \otimes \left( \hat{u} \ast w^{k+1} - \hat{d} \right) \right] \right\|}.$$
To avoid the direct computation of $\frac{\partial u}{\partial m}$, we apply the adjoint-state method (Plessix 2006).

The gradient with respect to the model $m$ can be obtained by cross-correlating the forward-propagated wavefield with the backward-propagated wavefield excited by the adjoint source located at the receivers. The adjoint source is written as

$$r = \frac{1}{\|u\|} \left[ w^{k+1} \otimes (\hat{u} \ast w^{k+1} - \hat{d}) - \hat{u} \left[ w^{k+1} \otimes (\hat{u} \ast w^{k+1} - \hat{d}) \right] \right].$$  \hspace{1cm} (16)

This adjoint source has a main difference with the conventional FWI approach and this difference is given by the term $w^{k+1} \otimes (\hat{u} \ast w^{k+1} - \hat{d})$. The matching filter can push the predicted data closer to the observed data through convolution. In principal, the distance between these two data sets should be less than a half of a cycle to alleviate cycle skipping.

The cross-correlation, the adjoint operator of convolution, is used to complete adjoint operation. The above computations are performed trace by trace. The optimization problem based on the penalty function expands our search space to include both $w$ and $m$, which allows for a linearization in the model update process corresponding to the updated filter. Thus, the nonlinearity of the inversion is mitigated.

**NUMERICAL EXAMPLES**

In this section, we first analyse the convergence property of the proposed objective function using a modified Marmousi model. Here, we show the behavior of the objective function with different penalty parameters, which can give us some insights into selecting the parameter. We also compare the objective functions with and without normalization.
We then use the synthetic data generated from the modified Marmousi model to demonstrate the effectiveness of the proposed inversion approach. Finally, we apply our approach to a marine real data set from the offshore Western Australia.

**Behavior of the objective function**

As a start, we test the behavior of the proposed objective function using the modified Marmousi model shown in Figure 1a. The model size is 7.8 km horizontally and 1.8 km vertically. We apply the finite difference method with second order in time and twelfth order in space approximation to simulate the seismic wave propagation. The spatial and time intervals for the wavefield are 20 m and 2 ms, respectively. We also employ the absorbing boundary condition on all sides of the model. The source wavelet is a Ricker wavelet with a peak frequency of 6 Hz after filtering out the low frequencies below 3 Hz. We place 78 sources and 391 receivers evenly distributed on the surface to excite and record the seismic data.

The penalty parameter and the normalization term play an important role in the success of our inversion method. To provide more insights in the inversion method, we will illustrate the effects of the penalty parameter and the normalization on the objective function. The curves of the reduced objective functions \( J_{\lambda}(m, w(m, \lambda)) \) over \( m = (1 + a)m_{true} \) are plotted in Figure 2a. \( m_{true} \) indicates the true model (Figure 1a). \( m \) is used to produce the modeled data \( u(m) \). The velocity error \( a \) varies from -0.5 to 0.5.
When \( a = 0 \), the model \( \mathbf{m} \) is correct, and thus, the objective function reduces to global minimum at \( a = 0 \). When \( \lambda \) is very large \((10^6 \text{ and } 10^8)\), the data-misfit term has a small contribution to the objective function. Therefore, the matching filter can’t improve the data matching effectively, and the inversion will include local minima. When \( \lambda \) is very small \((1 \text{ and } 10^2)\), the ill-posed inversion problem for the matching filter can’t be conditioned well, thus the matching filter fails to represent the data misfit. Compared with the objective function of conventional FWI (black line in Figure 2a), our objective function with reasonable \( \lambda \) \((10^4, 10^7)\) has a larger convergence region. For comparison, we also plot the objective function without normalization in Figure 2b. The convergence region is also larger than L2-norm objective function, but the convexity and convergence are worse than that of the normalized objective function represented by \( J_\lambda \).

**Synthetic Marmousi model**

We then use the modified Marmousi model (Figure 1a) to illustrate the effectiveness of the proposed inversion algorithm. The parameter settings for the acquisition and model are consistent with that used in testing the behavior of the objective function in the previous section. We simulate the observed data using the true model with a constant density acoustic modeling operator. 78 shots and 391 receivers are evenly located at the surface of the model. The maximum offset is 7.8 km. The source wavelet is a 6-Hz Ricker wavelet without frequencies below 3 Hz. We start the inversion process using the initial model (Figure 1b), in which the velocity linearly increasing below the constant-velocity layer. The
inversion result for conventional L2-norm FWI is shown in Figure 3a. Due to the inaccurate initial model and the lack of low frequencies, the inversion is trapped into a local minimum and the inverted velocity in the left part is contaminated with artifacts and far from the true model.

In the proposed inversion approach, we set the penalty parameter as $10^4$. The length of the matching filter is as long as the observed data. We, first, solve for the matching filter by 20 conjugate gradient (CG) iterations using the initial model, the data misfits decrease to about 50%. Figure 4a shows the inverted matching filter for a source located at $x_s = 4.04$ km for the initial model. We can see that the matching filter has been computed stably even with multiple arrivals. The matching filter is far away from the objective of an approximate zero-lag Dirac delta function, especially for the far-offset traces. The matching filter indicates the misfits between the modeled and observed data, and thus we can use the matching filter to transform the modeled data towards the observed data. In this case, the cycle skipping will be mitigated. Given the inverted matching filter, we can compute the gradient with respect to the velocity model. We resort to the backtracking line-search approach to estimate the step length, which satisfies the Wolfe condition (Wolfe 1969). The inversion converges to a reasonable result after 200 iterations, shown in Figure 3b, starting from the linearly increasing model (Figure 1b). To give you a general picture of the inversion accuracy, the differences between the true velocity model shown in Figure 1a and the inverted velocity models shown in Figures 3a and 3b are shown in Figures 3c and 3d,
respectively. We can see that the inverted velocity model using the proposed method is closer to the true model than that using the conventional FWI especially for the left part. Figure 4b shows the inverted matching filter for a source located at $x_s = 4.04$ km. We can see that the matching filter for the inverted model is well focused near zero lag. Figure 5 shows the convergence history of the normalized data misfits, i.e., $\| \hat{u}(m) - \hat{d} \|^2$ and the L2 norm of penalized normalized matching filter, i.e., $\| T \hat{w} \|^2$. Both of them drop to a relatively small value.

For a detailed comparison, we also plot in Figure 6 the vertical profiles of the true, initial and inverted velocity models at distance 2.5 and 5 km. The comparison of these vertical profiles further proves that the proposed inversion method estimates the subsurface model more accurately than the conventional FWI. At last, we compare the shot gathers at $x_s = 6.04$ km in Figure 7. As shown in Figures 7a and 7c, the shot gather for the initial model has a considerable deviation from the observed data, especially at far offsets. Figures 7b and 7d show that the shot gather for the inverted model (Figure 3b) match the observed data very well.

**Field data example**

The field data used in this example come from offshore Western Australia, acquired by courtesy of CGG with a variable depth streamer. The raw data set has 1824 shots with about 18.75 m horizontal interval. We only use 116 shots with about 93.75 m shot interval in our inversion to
reduce the computational burden. Every shot gather is recorded by a streamer, which contains 648 receivers with a 12.5 m horizontal interval. The minimum and maximum offsets are 0.169 km and 8.256 km, respectively. The maximum recording time is about 7s. Figure 8 shows the original shot gather from the source at a distance of 4.5 km. The minimum available frequency in the data is close to 2.5 Hz, but the low-frequency components are contaminated with strong noise.

We first apply a low-pass filter to the provided data set. The filtered shot gather with the source at $x_s=4.5$ km is shown in Figure 9a, and correspondingly its frequency spectrum is shown in Figure 9b. We estimate the source wavelet by inverting the near-offset early arrival data, and Figure 10 shows the estimated wavelet corresponding to the provided field data. The initial model is shown in Figure 11a, which is constructed from a brute-stack time-domain velocity analysis approach (Kalita and Alkhalifah 2019).

We set the penalty parameter used in the proposed inversion algorithm to $10^4$, which is same as the synthetic example. The matching filter is optimized with 20 CG iterations before every update of velocity model. Starting from the initial model shown in Figure 11a, we update the velocity model using the proposed method with 100 iterations and the inversion result is shown in Figure 11b. We can see that the long-wavelength structures are recovered well. The convergence history of the normalized data misfits is shown in Figure 12. We can see that the data misfits have been reduced by more than 50%, even though the observed data include obvious noise. Subsequently, we implement a multi-scale FWI including normalization to gradually introduce higher resolution structures. During this inversion process, the high frequencies are added gradually by setting the maximum frequency allowed by low-pass filter to 5, 6, 8, 10, 12, 15 and 20 sequentially. Figure 11c shows the final velocity model estimated by a multi-scale inversion process. For comparison, we also
perform conventional FWI on the low-pass filtered field data shown in Figure 9. The inversion result is shown in Figure 11d. We can see that a low-velocity zone at $x = 10.0$ km is retrieved in the shallow part, which is probably one of the inversion artifacts arising from cycle skipping. The accuracy of the inversion results will be evaluated based on data misfit and image quality in following.

We compare the shot gather and the matching filter at a distance of 4.5 km to examine the inverted velocity generated by the proposed method. Figure 13a shows the comparison between the modeled data for the initial velocity (Figure 11a) and the observed data (Figure 9a). Figures 13b and 13c show the shot gather comparison for the inverted velocity using the proposed method (Figure 11b) and the conventional FWI (Figure 11d), respectively. We can see that the initial velocity and conventional FWI result fail to predict the kinematical information of the diving waves especially for the far offsets. The cycle skipping occurs with conventional FWI because of the large data error, whereas the proposed inversion algorithm renders the seismic data fitting well. As shown in Figure 13b, both the diving waves and reflections are generated with relatively accurate kinematical information. The matching filters for the initial and inverted velocities are shown in Figures 14a and 14b, respectively. The matching filter for the inverted velocity is better focused near zero lag than the initial one, which also verifies the effectiveness of the proposed method. The vertical strips shown in the matching filters arise from the strong noise that contaminates the low-pass filtered shot gathers (Figure 9a). An appropriate pre-processing of the seismic data can help suppress these strips.

We apply reverse time migration (RTM) using the four velocity models shown in Figure 11. The corresponding images are shown in Figure 15. We can see that the estimated velocity model using the proposed method improves the image quality considerably. Compared with the RTM image shown in Figure 15d, the reflectors in Figure 15b are more continuous and better focused.
especially for the reflectors in the distance range $4\text{km}<x<6\text{km}$. The image using the final FWI result, shown in Figure 15c, indicates the structures in the investigated area with higher resolution.

**DISCUSSION**

In this study, we design a new objective function for data-domain extended waveform inversion by introducing a nonphysical extension, i.e., the matching filter for every trace. The proposed objective function consists of a data-misfit term and a penalty term. The balance between these two terms is controlled by a penalty parameter, which influences the convexity of the objective function. We normalize the two data sets rather than the matching filter to maintain the linearity of the inner loop corresponding to the deconvolution. For one thing, the normalization can suppress the amplitude trade-off between the matching filter and the modeled data, thus avoiding satisfying the objective by simply decreasing the coefficients of filter. For another, it alleviates the dependence of the objective function on accurately simulating the amplitudes, crucial in the real case. In short, the normalization of the data can improve the convexity and convergence properties of the objective function.

A reasonable evaluation of the matching filter is critical for the success of our inversion method. The inner loop for the matching filter is a 1-D linear inverse problem. We stably update the matching filter using a CG iterative method. The iterative process for the sub-problem of optimizing the filter increases the computational cost by about 30% compared to the conventional inversion based on the L2-norm objective function.

Regularization is an essential element in deconvolution in order to suppress instable solutions arising from frequencies with amplitudes close to zero. In our examples, the
penalty parameter $\lambda$ in the regularized deconvolution is determined by comparing the behavior of the objective function with different values of $\lambda$. L-curve-based method is also a potential way to choose a proper penalty parameter (Hansen and O’Leary 1993). These numerical examples suggest that a single fixed value of $\lambda$ is generally sufficient. When the penalty parameter $\lambda$ is too small or large, the sub-problem for inverting for the matching filter shows serious ill-posedness. Therefore, the convergence domain of the objective function doesn’t shrink as we increase the penalty parameter as some other extended objective functions using the subsurface offset image or the reconstructed wavefield. The inner problem for the extended images or the reconstructed wavefield is far more expensive than the deconvolution. In terms of practicality, the proposed method can provide a reasonable model for conventional FWI.

We use high-frequency information in FWI (the field data example) to delineate the high-resolution structures in the subsurface model. The high-frequency FWI further leads to high-resolution images. However, the FWI result still has the risk of data overfitting, which can be alleviated by a proper regularization technique (Asnaashari et al. 2013; Li et al. 2020). The matching filter-based extended inversion algorithm can be easily applied to elastic or anisotropic FWI, which usually suffers from more serious ill-posedness and uncertainty because of the extra multi-parameters trade-off challenges (Vigh et al. 2014; Huang et al. 2020). We can also combine this inversion method with the regularization techniques to guide the inversion to recover the subsurface model with higher resolution and accuracy. We will explore more effective regularization methods to alleviate the ill-posedness
problem in future work.

CONCLUSION

We proposed a normalized penalty objective function for full waveform inversion to exploit an extended search space and to avoid cycle skipping. The penalty function is optimized over the extended model: the matching filter and the subsurface model. The matching filter, as an unphysical extension, can force the predicted data towards the observed data and improve the data fitting. We can invert for the matching filter stably by solving a linear least-square problem iteratively. The normalization of the data plays an important role in improving the convexity and convergence properties of the objective function. The numerical examples show that the matching filter-based extended FWI can recover the velocity reasonably when conventional FWI suffers from cycle skipping.

APPENDIX

To derive the gradient of the objective function $J$ with respect to the model parameter $\mathbf{m}$, where

$$J = \frac{1}{2} \left\| \mathbf{u}(\mathbf{m}) * \mathbf{w}^{k+1} - \mathbf{d} \right\|^2,$$

we compute the derivative of the objective function $J$ with respect to $\mathbf{m}$:

$$\frac{\partial J}{\partial \mathbf{m}} = \left\langle \frac{\partial \mathbf{u}}{\partial \mathbf{m}}, \mathbf{w}^{k+1} \otimes \left(\mathbf{u} * \mathbf{w}^{k+1} - \mathbf{d}\right) \right\rangle.$$

(A-2)
To simplify the derivation, we use $\gamma$ to represent $w^{k+1} \otimes (\hat{u} * w^{k+1} - \hat{d})$, that is:

$$\gamma = w^{k+1} \otimes (\hat{u} * w^{k+1} - \hat{d}).$$  \hfill (A-3)

The detailed derivation for the gradient is given by:

$$\frac{\partial J}{\partial m} = \left\langle \frac{\partial \hat{u}}{\partial m}, \gamma \right\rangle$$

$$= \frac{\partial}{\partial m} \left( \int u(x,t)\gamma(x,t) \, dt \right)$$

$$= \frac{\partial m}{\partial m} \int \gamma(x,t) \frac{\partial u(x,t)}{\partial m} \, dt - \frac{\partial m}{\partial m} \left( \int u(x,t)\gamma(x,t) \, dt \right) \frac{\partial u(x,t)}{\partial m} \, dt$$

$$= \frac{1}{\|u\|} \left\langle \frac{\partial u}{\partial m}, \gamma \right\rangle - \frac{1}{\|u\|} \left\langle \frac{\partial u}{\partial m}, \hat{u} \cdot (\hat{u} \cdot \gamma) \right\rangle$$

$$= \left\langle \frac{\partial u}{\partial m}, \frac{1}{\|u\|} \gamma - \hat{u} \hat{u} \cdot \gamma \right\rangle$$

Substitute (A-3) into the gradient formula of (A-4), the final expression for the gradient is given by:

$$g_m = \frac{\partial J}{\partial m} = \left\langle \frac{\partial u}{\partial m}, \frac{1}{\|u\|} \left( w^{k+1} \otimes (\hat{u} * w^{k+1} - \hat{d}) - \hat{u} \left[ w^{k+1} \otimes (\hat{u} * w^{k+1} - \hat{d}) \right] \right) \right\rangle.$$ \hfill (A-5)
List of Figure Legends

1 (a) The modified Marmousi model; (b) The initial velocity model.
The objective function using different penalty values $\lambda$ over velocity error scaling of the Marmousi model: (a) the proposed one with normalization and (b) that without normalization.
The inverted velocity models using (a) conventional FWI and (b) our method and (c-d) the differences between the true velocity model shown in Figure 1a and the inverted velocity models shown in Figures 3a and 3b.
The matching filter for a source at $x_s = 4.04$ km using (a) the initial model (Figure 1b) and (b) the inverted model (Figure 3b).
Convergence history of (a) the normalized data misfits, i.e., $\|\hat{u}(m) - \hat{d}\|^2$ and (b) the L2 norm of penalized normalized matching filter, i.e., $\|\hat{T}\hat{w}\|^2$. 

This article is protected by copyright. All rights reserved.
Comparison of the vertical profiles at distance (a) 2.5 km and (b) 5 km. Red line: the true model shown in Figure 1a; cyan line: the initial model shown in Figure 1b; blue line: the matching filter-based extended FWI result shown in Figure 3b; green line: the conventional FWI result shown in Figure 3a.
Comparison of the shot gathers at \( x_s = 6.04 \, \text{km} \) generated by (a) the initial model (left panel) and (b) the inverted model (left panel) with the observed data (right panel in (a) and (b)), (c) the data difference between the two shot gathers shown in (a) and (d) the data difference between the two shot gathers shown in (b).
A raw shot gather from the field data for a source at a distance of 4.5 km.
9 (a) A low-pass filtered shot gather for a source located at distance 4.5 km and (b) its frequency spectrum. The low-pass filtered shot gathers are used to test the proposed inversion method.
The source wavelet estimated from the provided field data.
11 (a) The initial velocity, (b) the inverted velocity with the proposed inversion method, (c) the multi-scale FWI result starting from the inverted velocity shown in (b), (d) conventional FWI result using the same input data as that of (b) for the field data example.
Convergence history of the normalized data misfits, i.e., \( \| \hat{u}(m) - \hat{d} \| \), for the field data example.

![Graph showing the normalized data misfits over iterations.](image-url)
Comparison of the shot gathers at $x_s = 4.5$ km in the field data example for (a) the initial velocity, (b) the inverted velocity using the proposed method (Figure 11b) and (c) the inverted velocity using the conventional FWI (Figure 11d). The modeled data is shown in the side panel, and the observed data is shown in the middle panel.
Comparison of the matching filters at $x_s = 4.5$ km in the field data example for (a) the initial velocity (Figure 11a) and (b) the inverted velocity (Figure 11b).
15 RTM images in the field data example using (a) the initial velocity in Figure 11a, (b) the inverted velocity in Figure 11b, (c) the final multi-scale FWI result in Figure 11c and (d) RTM image using the conventional FWI result shown in Figure 11d.
REFERENCES


This article is protected by copyright. All rights reserved.


Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.