A nonlinear scheme to perform linearized waveform inversion with velocity updating

Alejandro Cabrales-Vargas, and Rahul Sarkar

ABSTRACT

In this paper we change the original proposal for linearized waveform inversion with velocity updating into an iterative nonlinear scheme. The previous linear scheme resulted in an objective function, in which the term responsible for maximization of the image power was unbounded along all directions of perturbations of the background velocity model. In the new scheme that we experiment with in this paper, this problem is not present. One important difference of the new method as compared to the original is that in the original method, the goal was to solve a single linear inverse problem to estimate both the perturbations in the background and reflectivity models, but in the new method we solve a sequence of linear inverse problems. We first present an analysis of the previous method identifying the problem, and then present the details of the new iterative scheme that we are experimenting with to overcome it. It should be noted that we are in the beginning stages of development of the method, and it will likely change in the future as we develop a better understanding of how it is performing. We present some numerical experiments to support our claims that provide insight into the properties of the objective function.

INTRODUCTION

Accurate estimation of the reflectivity of the subsurface has been a continuous endeavor in oil and gas exploration. Kirchhoff migration estimates the reflectivity using ray theory, but it breaks down in the presence of multipathing. In comparison two-way wave equation based methods, like reverse-time migration (RTM) (Kosloff and Baysal, 1983; Baysal et al., 1983; Gazdag and Carrizo, 1986) and linearized waveform inversion (LWI), *a.k.a.* least-squares migration (Nemeth and Schuster, 1999; Duquet et al., 2000; Dai et al., 2013; Wong et al., 2015; Schuster, 2017), are more robust in the presence of multipathing caused by complex geology, and produce better images of the subsurface.

During the LWI process the assumption is that the background model has been estimated accurately, and the goal is to recover the high wavenumber reflectivity model. The background model is never changed during this process. An important fact about LWI is that incorrect background models will generally produce incorrect reflectivity

models, because Born linearization is not accurate. In addition, it has also been reported that the convergence rate of LWI slows down in the presence of an incorrect background model (Luo and Hale, 2014). In some cases and depending on the application, one may or may not need a very accurate estimate of the background model. For instance if one is interested in exploration at the regional or prospect scale, one only requires the background model to be precise to the extent that it honors the kinematics correctly, and is able to correctly position the seismic reflectors. However once a reservoir has been discovered, an important goal is to perform careful interpretation and AVO studies for volumetric assessment. In such cases, small inaccuracies in the background model can mislead the interpretation, especially if subtle features indicate important lithologic and/or petrophysical variations, and thus one needs an accurate estimate of the background model in these scenarios.

Recent research has thus focused on improving LWI in the presence of velocity errors (Yang et al., 2018). These challenges also motivated the original conception of the technique *Linearized Waveform Inversion with Velocity Updating (LWIVU)* (Cabrales-Vargas, 2018b). In this report we continue our work on LWIVU, which attempts to simultaneously invert for both the reflectivity and the background components of the subsurface model. However we discovered a problem with the original objective function, specifically in the term intended to promote solutions that maximize the power of the migration image. In principle, the problem can be avoided by using differential semblance optimization (DSO) in the subsurface offset domain (Cabrales-Vargas, 2018a), but the additional cost due to the extension is considerable, especially in 3D. Moreover the DSO operator can be difficult to stabilize. In this paper we explore an idea to fix the issue with the previous LWIVU objective function, without going into the extended domain.

This paper is structured as follows. We start by reviewing the original ideas and motivation behind LWIVU, and briefly discuss the formulation of the problem as a nonlinear optimization problem. We then discuss the linearization of the problem that was originally performed to derive the LWIVU objective function as discussed in Cabrales-Vargas (2018b). Additionally we present some mathematical analysis that shows why this objective function does not have the desired properties. Next, we introduce the iterative LWIVU algorithm that we are currently experimenting with to overcome this problem. Finally, we present some results from numerical experiments that provide evidence to support our claims.

THEORY

We first introduce some notation that we will use in this paper. Let us assume that the true velocity model $\mathbf{v}^{\text{true}}$ admits scale separation as a sum of a low wavenumber component (the *true background model*), which we will denote as $\mathbf{b}^{\text{true}}$, and a high wavenumber component (the *true reflectivity model*), which we will denote as $\mathbf{r}^{\text{true}}$, and so the following equation holds

$$\mathbf{v}^{\text{true}} = \mathbf{b}^{\text{true}} + \mathbf{r}^{\text{true}}. \tag{1}$$

The above assumption is widespread in seismic imaging, and is one of the key assumptions of seismic migration algorithms, including LWI. In everything that follows, we will assume that the data that we migrate is the Born linearized data, which we will denote as $\mathbf{d}^{\text{true}}$. Using the notation $\mathbf{L}(\mathbf{b})$ to denote the Born linearization operator at background model $\mathbf{b}$, we thus have

$$\mathbf{d}^{\text{true}} = \mathbf{L}(\mathbf{b}^{\text{true}})\mathbf{r}^{\text{true}}. \tag{2}$$

The migration image with some background model $\mathbf{b}$ will be denoted as $\mathbf{I}(\mathbf{b})$, and concretely it is obtained as the adjoint of the Born linearization operator applied to the true Born data $\mathbf{d}^{\text{true}}$, i.e. the following equation holds

$$\mathbf{I}(\mathbf{b}) = \mathbf{L}^{\mathbf{T}}(\mathbf{b})\mathbf{d}^{\text{true}}. \tag{3}$$

The full-waveform inversion Gauss-Newton Hessian is similarly defined for a background model $\mathbf{b}$ as $\mathbf{H}(\mathbf{b}) = \mathbf{L}^{\mathbf{T}}(\mathbf{b})\mathbf{L}(\mathbf{b})$, while the wave equation migration velocity analysis (WEMVA) operator, denoted as $\mathbf{W}(\mathbf{b})$, is defined as the derivative of the migration image with respect to the background, that is

$$\mathbf{W}(\mathbf{b}) = \frac{\partial \mathbf{I}}{\partial \mathbf{b}}(\mathbf{b}). \tag{4}$$

The tomographic operator, denoted as $\mathbf{T}(\mathbf{b})$, constitutes the derivative of the Born data with respect to the background model (Almomin, 2013; Barnier and Almomin, 2014),

$$\mathbf{T}(\mathbf{b}) = \frac{\partial \mathbf{d}}{\partial \mathbf{b}}(\mathbf{b}) = \frac{\partial \mathbf{L}}{\partial \mathbf{b}}(\mathbf{b}, \mathbf{r}). \tag{5}$$

Note that the derivative of the Born modeling operator with respect to the background model is not a matrix operator by itself. It is its action upon the reflectivity what becomes a matrix operator.

Motivation behind LWIVU

The original formulation of LWIVU was founded on two (approximate) principles which are described below.

1. *Maximization of migration image power*: It is believed in the seismic imaging community that the maximum focusing of the migration image takes place when

it is evaluated at the true background model $\mathbf{b}^{\text{true}}$, although no mathematical proof of this fact is currently known to exist to the authors. This condition is expressed by saying that $\mathbf{b}^{\text{true}}$ maximizes the migration image power, i.e. $\|\mathbf{I}(\mathbf{b})\|_2^2$ attains the maximum at $\mathbf{b}^{\text{true}}$. In fact we can positively say in this paper that this assertion is false in general, as shown by numerical experiments later, where random one dimensional scans of $\|\mathbf{I}(\mathbf{b})\|_2^2$ show that $\mathbf{b}^{\text{true}}$ is not a stationary point (see Figures 4a, 4b). The figures also show that $\|\mathbf{I}(\mathbf{b})\|_2^2$ restricted to random lines (at least in all the cases tried) through $\mathbf{b}^{\text{true}}$ have a maxima "close" to $\mathbf{b}^{\text{true}}$. Because of this reason we conclude that the principle of maximization of migration image power is a good heuristic to use to recover the background model, even though the actual statement is manifestly false in the absence of extra conditions. It should be noted that an analogous principle is believed to be true for extended migrated images in the extended subsurface offset domain (but it is not equivalent), which states that if the background model is correct then the energy is focused at zero subsurface offset, and if it is incorrect, then some energy is smeared at non-zero subsurface offsets.

We will also assume that an estimate of the true background model is known, which we will denote as $\mathbf{b_0}$, and refer to it as the *starting background model*. The quantity $\mathbf{\Delta b}^{\text{true}} = \mathbf{b}^{\text{true}} - \mathbf{b_0}$ is what we will try to recover, and we will refer to it as the *true background model perturbation*. Thus equivalently we can express our goal as

$$\text{maximize} \quad \Phi_1(\mathbf{\Delta b}) = \|\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b})\|_2^2 , \tag{6}$$

where the maximization is over the background model perturbation $\mathbf{\Delta b}$.

2. *Fitting data with Born linearization*: As (6) does not depend on the reflectivity, it is not possible to update $\mathbf{r}$ based on the objective function $\Phi_1(\mathbf{\Delta b})$ alone. For this purpose we will use a different criterion, namely that the data can be modeled using Born linearization. Therefore, we can invert for the reflectivity using LWI (for a fixed background model $\mathbf{b}$) in the data domain by minimizing the objective function

$$\left\|\mathbf{L}(\mathbf{b})\mathbf{r} - \mathbf{d}^{\text{true}}\right\|_2^2 , \tag{7}$$

or equivalently solve the corresponding normal equations (which in principle can be solved exactly),

$$\mathbf{L}(\mathbf{b})^{\mathbf{T}}\mathbf{L}(\mathbf{b})\mathbf{r} = \mathbf{L}(\mathbf{b})^{\mathbf{T}}\mathbf{d}^{\text{true}}. \tag{8}$$

Using notations already introduced this can be simply rewritten as

$$\mathbf{H}(\mathbf{b})\mathbf{r} = \mathbf{I}(\mathbf{b}). \tag{9}$$

Equation (9) can be cast into the optimization of the following objective function

$$\|\mathbf{H}(\mathbf{b})\mathbf{r} - \mathbf{I}(\mathbf{b})\|_2^2 , \tag{10}$$

which is known in the seismic imaging community as *"LWI in the model domain"*. To solve this problem, one can precompute the approximate Hessian by means of point-spread functions (PSF) (Fletcher et al., 2016). The objective functions (7) and (10) assume that the background model remains fixed during the inversion. However we can choose to consider $\mathbf{b}$ as an additional parameter in this inversion. Hence with $\mathbf{b}_0$ as an estimate of the background model we have the second LWIVU goal

$$\text{minimize} \quad \Phi_2(\boldsymbol{\Delta}\mathbf{b}, \mathbf{r}) = \|\mathbf{H}(\mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b})\mathbf{r} - \mathbf{I}(\mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b})\|_2^2 . \tag{11}$$

Original formulation of LWIVU

In this section we briefly discuss how the LWIVU goals were originally combined into a single objective function, and what we have discovered recently to be a key error in the procedure followed.

Derivation of the original method

The two LWIVU goals can be combined into a minimization problem of a single objective function

$$\Phi(\mathbf{r}, \boldsymbol{\Delta}\mathbf{b}) = \|\mathbf{H}(\mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b})\mathbf{r} - \mathbf{I}(\mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b})\|_2^2 - \lambda \|\mathbf{I}(\mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b})\|_2^2 , \tag{12}$$

where λ is a trade-off parameter, and the minus sign expresses the fact that the second term is being maximized. The second term, which corresponds to the objective function (6), represents a nonlinear relationship between the migration image and the perturbation in the background. In the previous implementation of LWIVU (Cabrales-Vargas, 2018b), the image was first linearized as

$$\mathbf{I}(\mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b}) = \mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b} + O(||\boldsymbol{\Delta}\mathbf{b}||^2), \tag{13}$$

which constitutes a first order approximation when $\boldsymbol{\Delta}\mathbf{b}$ is small. Hence, with this approximation, i.e. neglecting the second order terms in the image, the objective function in (12) becomes

$$\Phi(\mathbf{r}, \boldsymbol{\Delta}\mathbf{b}) = \|\mathbf{H}(\mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b})\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b} - \mathbf{I}(\mathbf{b}_0)\|_2^2 - \lambda \|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b}\|_2^2 . \tag{14}$$

Since recalculating the Hessian is extremely costly, another approximation was used in the regime $||\mathbf{\Delta b}||_2 << ||\mathbf{b}_0||_2$, namely $\mathbf{H}(\mathbf{b}_0 + \mathbf{\Delta b}) \approx \mathbf{H}(\mathbf{b}_0)$, which yields the original LWIVU objective function

$$\Phi(\mathbf{r}, \mathbf{\Delta b}) = \|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} - \mathbf{I}(\mathbf{b}_0)\|_2^2 - \lambda \|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}\|_2^2 . \tag{15}$$

The key feature that was attractive about the above formulation (15) is that it is a linear least squares minimization problem, and one can solve it using existing algorithms like the conjugate gradient method (Cabrales-Vargas, 2018a).

Problem with the original formulation

Even though the original LWIVU problem is easy to solve (for a particular choice of λ), the objective function (15) has a serious drawback, which we will now describe. The key point is that first performing the linearization of $\mathbf{I}(\mathbf{b})$ and plugging the expression into $||\cdot||_2^2$, is not the same as the linearization of $||\mathbf{I}(\mathbf{b})||_2^2$. In fact doing the former completely changes the properties of $||\mathbf{I}(\mathbf{b})||_2^2$. Let us include another term in the expansion of the migration image

$$\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b}) = \mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} + \mathbf{Z}(\mathbf{b}_0, \mathbf{\Delta b}) + O(||\mathbf{\Delta b}||^3), \tag{16}$$

where $\mathbf{Z}(\mathbf{b}_0, \mathbf{\Delta b})$ incorporates all the second order terms in the expansion. Plugging (16) into the function $||\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b})||_2^2$ gives

$$\begin{aligned} ||\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b})||_2^2 &= \mathbf{I}(\mathbf{b}_0)^\mathbf{T}\mathbf{I}(\mathbf{b}_0) + 2\mathbf{\Delta b}^\mathbf{T}\mathbf{W}(\mathbf{b}_0)^\mathbf{T}\mathbf{I}(\mathbf{b}_0) \\ &\quad + 2\mathbf{Z}(\mathbf{b}_0, \mathbf{\Delta b})^\mathbf{T}\mathbf{I}(\mathbf{b}_0) + \mathbf{\Delta b}^\mathbf{T}\mathbf{W}(\mathbf{b}_0)^\mathbf{T}\mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} + O(||\mathbf{\Delta b}||^3) \\ &= \|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}\|_2^2 + 2\mathbf{Z}(\mathbf{b}_0, \mathbf{\Delta b})^\mathbf{T}\mathbf{I}(\mathbf{b}_0) + O(||\mathbf{\Delta b}||^3). \end{aligned} \tag{17}$$

Notice the additional term $2\mathbf{Z}(\mathbf{b}_0, \mathbf{\Delta b})^\mathbf{T}\mathbf{I}(\mathbf{b}_0)$, which is second order and is not accounted for by the second term in the LWIVU objective function (15). Thus in particular the approximation of $\Phi_1(\mathbf{\Delta b})$ is not order accurate, which is problematic in this case as the curvature of the function is changed near $\mathbf{\Delta b}^{\text{true}}$ (in fact it becomes opposite). This is because while we expect $\Phi_1(\mathbf{\Delta b})$ to be upper bounded, the approximation $\|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}\|_2^2$ is not, and is actually a convex function with a constant Hessian that is at least positive semidefinite. As consequence, there is no upper bound or maximum close to the correct background model perturbation. We provide some numerical results to demonstrate this effect in the next section (see Figures 3a, 3b). This implies that one can move along any arbitrary direction with strictly positive curvature, and increase the objective function $\|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}\|_2^2$ indefinitely, which is clearly not what is desired.

Iterative nonlinear scheme for LWIVU

Given the shortcoming of the objective function (15) we now propose to instead minimize the full nonlinear objective function (14). However as a first try, we will continue to work under the assumption $||\mathbf{\Delta b}||_2 << ||\mathbf{b}_0||_2$, so that the approximation $\mathbf{H}(\mathbf{b}_0 + \mathbf{\Delta b}) \approx \mathbf{H}(\mathbf{b}_0)$ is accurate. An easy technical result formalizing the regime where such an approximation is valid is provided in Appendix B. So the objective function that we will minimize is

$$\Phi(\mathbf{r}, \mathbf{\Delta b}) = \|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} - \mathbf{I}(\mathbf{b}_0)\|_2^2 - \lambda \|\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b})\|_2^2 . \tag{18}$$

This makes the first term a linear least squares problem over both $\mathbf{\Delta b}$ and $\mathbf{r}$, while the second term is nonlinear and non-quadratic. Such problems can be solved (for local minima) using gradient-based methods such as the non-linear conjugate gradient algorithm (Nocedal and Wright, 2006). However as a first try we are going to try something simpler — the idea will be to linearize the second term at each step, so that the resulting problem is still convex, solve this problem, and then relinearize at the new point and continue this process until convergence. The pseudocode of the algorithm is outlined in Algorithm 1, and the details are provided in Appendix A. An attractive feature of this approach is that each subproblem is still a convex problem (in fact it is a quadratic problem with a Hessian that is semidefinite).

Algorithm 1 Iterative linearized waveform inversion with velocity updating

procedure ITERATIVE LWIVU ($\mathbf{b}_0$, n)
  Initialization: $\mathbf{r}_0 \leftarrow \mathbf{0}$; $\mathbf{\Delta b}_0 \leftarrow \mathbf{0}$
  for i = 0 to n **do**
    $\mathbf{f}_i \leftarrow \mathbf{I}(\mathbf{b}_0) - \mathbf{H}(\mathbf{b}_0)\mathbf{r}_i + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}_i$
    $\mathbf{g}_i \leftarrow \mathbf{W}(\mathbf{b}_0 + \mathbf{\Delta b}_i)^{\mathbf{T}}\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b}_i)$
    $(\mathbf{r}^*, \mathbf{\Delta b}^*) \leftarrow$ minimize $\|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} - \mathbf{f}_i\|_2^2 - \lambda \mathbf{g}_i^{\mathbf{T}}\mathbf{\Delta b}$
    $\mathbf{r}_{i+1} \leftarrow \mathbf{r}_i + \mathbf{r}^*$; $\mathbf{\Delta b}_{i+1} \leftarrow \mathbf{\Delta b}_i + \mathbf{\Delta b}^*$
  end for
end procedure

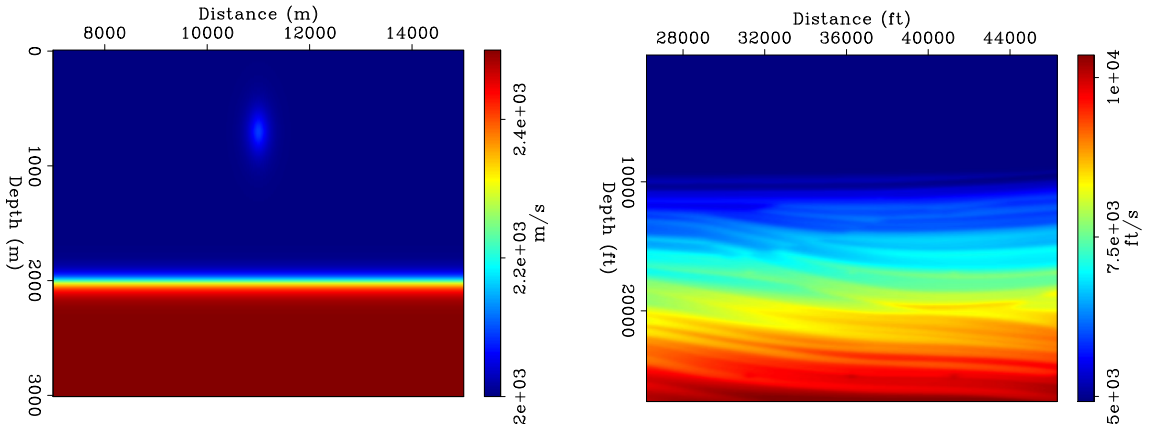
NUMERICAL RESULTS

In this section we present some numerical experiments to first illustrate the incorrect behavior of the image power maximization term, followed by one dimensional scans of the fully nonlinear image power maximization objective function. We also show the gradient of this function on a simple example that illustrates that it has good properties. Even though we show all the subsurface models in velocity, the actual computations were performed in the slowness squared domain.

Experiment setup

For the tests we used a flat-layer model and a portion of the Sigsbee model, shown in Figure 1. The flat-layer background model (Figure 1a) includes a positive Gaussian anomaly in the velocity, which corresponds to a negative anomaly in slowness squared. This model constitutes the true background model, $\mathbf{b}^{\text{true}} = \mathbf{b}_0 + \mathbf{\Delta b}^{\text{true}}$. The incorrect background model $\mathbf{b}_0$ does not contain the anomaly. Hence, this anomaly represents the true perturbation in the background $\mathbf{\Delta b}^{\text{true}}$. On the other hand, the true Sigsbee background model (Figure 1b) does not contain a Gaussian anomaly. However, to construct an incorrect background model we added a negative Gaussian anomaly in velocity to the true background model. In this case this negative anomaly in velocity corresponds to a positive anomaly in slowness squared.

We synthesized the true data $\mathbf{d}^{\text{true}}$ in both cases by applying Born modeling to the corresponding reflectivity models (Figure 2) using the true background models. For the flat-layer model the acquisition geometry consists of 101 shots regularly spaced every 140 m, and receivers every 20 m. For the Sigsbee model the acquisition geometry consists of 54 shots spaced every 500 ft, and receivers every 75 ft.

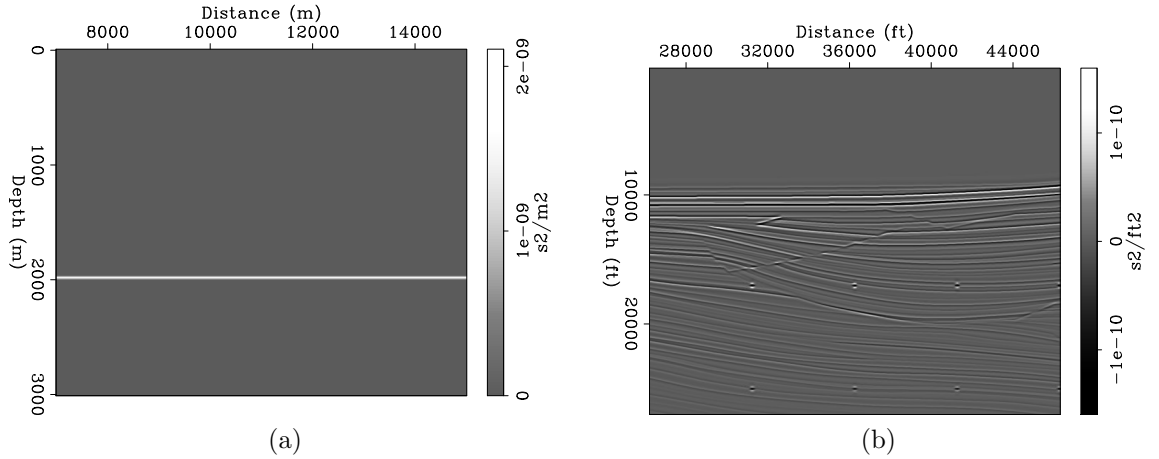


(a) (b)

Figure 1: Background velocity models in velocity. a) Flat model; b) Sigsbee model. [**ER**]

Testing the migration image power term in the original LWIVU objective function

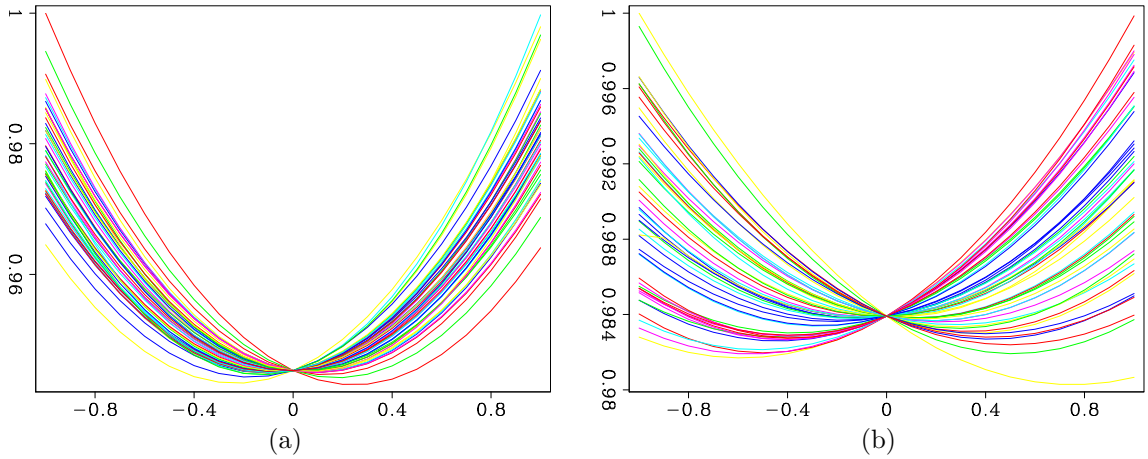
The first test constitutes the demonstration that the image power maximization term in the original LWIVU objective function (15) does not have the desired properties, namely it is in fact convex. To do this, we create trial perturbations in the background model, given as $\mathbf{\Delta b}^{\text{true}} + \alpha \mathbf{\Delta b}_{\text{rand}}$, where $\mathbf{\Delta b}^{\text{true}}$ is the true anomaly, $\mathbf{\Delta b}_{\text{rand}}$ represents randomly distributed perturbations whose amplitudes are confined within the interval $\big[-\max(\mathbf{\Delta b}^{\text{true}}), \max(\mathbf{\Delta b}^{\text{true}})\big]$, and α is a scalar parameter with values



(a) (b)

Figure 2: Reflectivity models in slowness squared. a) Flat model; b) Sigsbee model. [**ER**]

$-1, -0.9, ..., 0.9, 1$. Then the functional $\left\|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\left[\mathbf{\Delta b}^{\text{true}} + \alpha \mathbf{\Delta b}_{\text{rand}}\right]\right\|_2^2$ was evaluated using different random realizations of $\mathbf{\Delta b}_{\text{rand}}$, and exploring along each direction with the α values, so $\alpha = 0$ corresponds to the evaluation of the objective function using the true background anomaly.



(a) (b)

Figure 3: Experiment results of 50 random directions of the $\|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}^{\text{true}} + \alpha \mathbf{\Delta b}_{\text{rand}}\|_2^2$ functional in the a) Flat model, b) Sigsbee model. Each random direction was explored by means of the α scalar. [**NR**]

Figure 3 shows the result of 50 tests of the evaluation of the objective function using random directions, $\mathbf{\Delta b}_{\text{rand}}$, each one for the 21 values of α. Notice that there is no upper bound, and the minima of the curves are clustered at the vicinity of the true anomaly. These results prove that the this objective function is convex, as it was already mentioned before.

Test on the fully nonlinear migration image power term

We next perform similar tests for the fully nonlinear image power maximization objective function in equation (6). Again we evaluate this function along random directions. The results for both the models are shown in Figure 4. These results confirm that this objective function is concave along the directions explored. Of course these experiments do not tell us for sure if it is in fact concave, and instead we can merely conclude that this may be an approximate heuristic. However, even though the maximum values of the curves cluster around $\alpha = 0$, they are not exactly there.

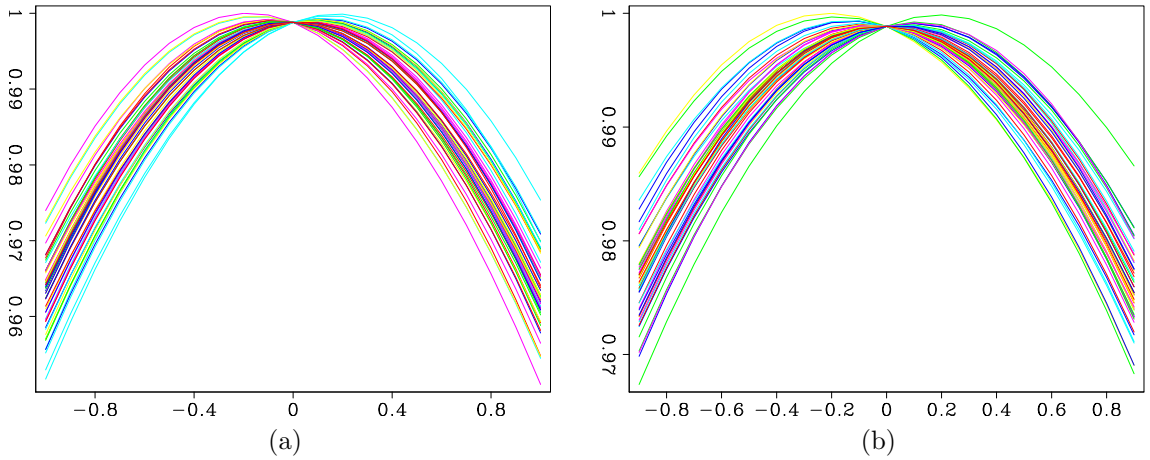


Figure 4: Experiment results of 50 random directions of the $\|\mathbf{I}(\mathbf{b}^{\text{true}} + \alpha\boldsymbol{\Delta}\mathbf{b}_{\text{rand}})\|_2^2$ functional in the a) Flat model, b) Sigsbee model. Each random direction was explored by means of the α scalar. [**NR**]

We finally show in Figure 5a the result of evaluating the gradient of the nonlinear term in equation (A-4) at the incorrect background, $\mathbf{b}_0$. For comparison, the true anomaly in slowness squared is shown in Figure 5b. This result indicates that the gradient points in the correct direction of the background model perturbation. However, the gradient also incorporates high wavenumber reflectivity components. We will address this problem in future implementations of the method using smoothing constraints.

DISCUSSION

The numerical tests demonstrate the mathematical fact that the original formulation of LWIVU did not have the expected behavior with respect to the image power maximization criteria. This encourages to look instead to try to solve the optimization problem treating the image power maximization term as a nonlinear term. The preliminary result of the gradient of this term for the flat model indicates that the inversion may proceed in the correct direction, although we know that we will need to enforce extra constraints (such as smoothing) for more complex models. For example, the result exhibits the presence of high wavenumber reflectivity amplitudes in the gradient. This issue will be addressed in the future.

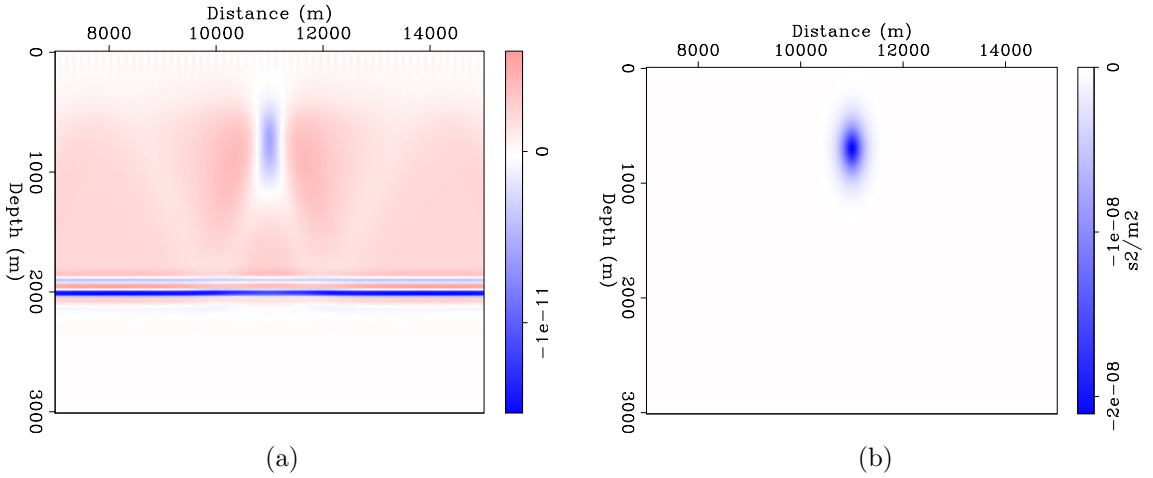


Figure 5: Flat model: a) First gradient of the nonlinear image power maximization term; b) True perturbation of the background in slowness squared. [**CR**]

We also mentioned before that the perturbation in the background model will not significantly change the result of the inversion with respect to the data fitting term. In other words, it is reasonable to work under the approximation that $\mathbf{H}(\mathbf{b}_0+\mathbf{\Delta b})\mathbf{r} \approx \mathbf{H}(\mathbf{b}_0)\mathbf{r}$. In Appendix B we offer a precise mathematical condition that quantifies this statement. However, there is also a heuristic justification to perform this approximation based on practical considerations. Firstly, the Hessian is seldom calculated exactly because of computational and storage limitations, and instead approximations such as the PSFs are used. This procedure is attractive because it only requires one pass of Born forward modeling followed by Born adjoint modeling to compute the impulse response of the Hessian. The reflectivity model consists of unit spikes appropriately scattered in the model space, placed to minimize the interaction of the computed PSFs. This procedure is cheaper in comparison to the more accurate procedure of applying the Born forward and adjoint operators independently to each grid point in the model space and storing the resulting PSFs.

We can construct elements of the Hessian by picking amplitude values of gridpoints from the PSF at lags measured from the center, where the original spike was seeded. Zero-lag points correspond to the diagonal of the Hessian. Off-diagonal elements correspond to different lags. Points corresponding to each lag are interpolated and smoothed in the model space. Therefore, this procedure yields a low-wavenumber estimation of the Hessian. On the contrary, we expect that the tomographic term in equation (A-15) supplies the action of the Hessian with comparatively high-wavenumber amplitude corrections, which would be filtered away after interpolation and smoothing. As a consequence, we can safely drop the tomographic term and approximate the Hessian as $\mathbf{H}(\mathbf{b}_0)$.

ACKNOWLEDGMENT

Alejandro Cabrales would like to thank Petróleos Mexicanos for the financial support. Both authors would like to thank the affiliates of the Stanford Exploration Project for financial support. The authors would also like to thank Guillaume Barnier for his comments on an early draft of this paper, that eventually led to the mathematical result in Appendix B.

APPENDIX A

In this appendix, we provide some of the missing details behind the derivation of Algorithm 1. We first explain how the expressions in the algorithm come up, and then show that the inner subproblem is indeed a quadratic function with a positive semidefinte Hessian.

Iterative linearization of the migration image power term

In the original expression in equation (18), suppose that before iteration i we are at a given point $(\mathbf{\Delta b}_i, \mathbf{r}_i)$, i.e. the estimate of the background model is $\mathbf{b}_0 + \mathbf{\Delta b}_i$, and the estimate of the reflectivity is $\mathbf{r}_i$. For the i^{th} iteration, our plan is to linearize the migration image power term around $(\mathbf{b}_0 + \mathbf{\Delta b}_i, \mathbf{r}_i)$, and then express the optimization problem in new coordinates obtained by translating the origin to $(\mathbf{\Delta b}_i, \mathbf{r}_i)$. With abuse of notation, and still using $(\mathbf{\Delta b}, \mathbf{r})$ to denote the new shifted coordinates, our objective function (18) becomes

$$\begin{aligned}\Phi(\mathbf{r}, \mathbf{\Delta b}) = &\left\|\mathbf{H}(\mathbf{b}_0)(\mathbf{r}_i + \mathbf{r}) - \mathbf{W}(\mathbf{b}_0)(\mathbf{\Delta b}_i + \mathbf{\Delta b}) - \mathbf{I}(\mathbf{b}_0)\right\|_2^2 \\ &- \lambda \left\|\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b}_i + \mathbf{\Delta b})\right\|_2^2 .\end{aligned} \tag{A-1}$$

We first collect all the terms that are fixed for the i^{th} iteration. Let us denote $\mathbf{f}_i = \mathbf{I}(\mathbf{b}_0) - \mathbf{H}(\mathbf{b}_0)\mathbf{r}_i + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}_i$, and $\mathbf{b}_i = \mathbf{b}_0 + \mathbf{\Delta b}_i$. Then we can rewrite (A-1) as

$$\Phi(\mathbf{r}, \mathbf{\Delta b}) = \left\|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} - \mathbf{f}_i\right\|_2^2 - \lambda \left\|\mathbf{I}(\mathbf{b}_i + \mathbf{\Delta b})\right\|_2^2 . \tag{A-2}$$

The first term in the right-hand side of equation (A-2) is a linear least squares problem. We linearize the second term around $(\mathbf{b}_0 + \mathbf{\Delta b}_i, \mathbf{r}_i)$ and obtain

$$\left\|\mathbf{I}(\mathbf{b}_i + \mathbf{\Delta b})\right\|_2^2 = \left\|\mathbf{I}(\mathbf{b}_i)\right\|_2^2 + (\mathbf{\Delta b})^{\mathbf{T}}\mathbf{W}(\mathbf{b}_i)^{\mathbf{T}}\mathbf{I}(\mathbf{b}_i). \tag{A-3}$$

It should be noted that $(\mathbf{\Delta b})^{\mathbf{T}}\mathbf{W}(\mathbf{b}_i)^{\mathbf{T}}\mathbf{I}(\mathbf{b}_i)$ is the gradient of the function $\left\|\mathbf{I}(\mathbf{b})\right\|_2^2$ at $\mathbf{b}_i$. Let us define

$$\mathbf{g}_i = \mathbf{W}(\mathbf{b}_i)^{\mathbf{T}}\mathbf{I}(\mathbf{b}_i), \tag{A-4}$$

and so we neglecting the constant term $\|\mathbf{I}(\mathbf{b}_i)\|_2^2$, we can express the approximation of (A-2) as

$$\Phi(\mathbf{r}, \boldsymbol{\Delta}\mathbf{b}) = \|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b} - \mathbf{f}_i\|_2^2 - \lambda(\boldsymbol{\Delta}\mathbf{b})^{\mathbf{T}}\mathbf{g}_i. \tag{A-5}$$

If $\mathbf{r}^*$ and $\boldsymbol{\Delta}\mathbf{b}^*$ are solutions to the minimization of (A-5), then the updated values of the background and the reflectivity models are obtained by undoing the translations, and so we have

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{r}^*, \quad \boldsymbol{\Delta}\mathbf{b}_{i+1} = \boldsymbol{\Delta}\mathbf{b}_i + \boldsymbol{\Delta}\mathbf{b}^*, \quad \mathbf{b}_{i+1} = \mathbf{b}_i + \boldsymbol{\Delta}\mathbf{b}^*. \tag{A-6}$$

Solving the subproblem

The objective function in equation (A-5) can be written after a variable change as

$$\begin{aligned}\Phi(\mathbf{r}, \boldsymbol{\Delta}\mathbf{b}) &= \|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b} - \mathbf{f}_i\|_2^2 - \lambda(\boldsymbol{\Delta}\mathbf{b})^{\mathbf{T}}\mathbf{g}_i \\ &= \left\| \begin{bmatrix}\mathbf{H}(\mathbf{b}_0) & -\mathbf{W}(\mathbf{b}_0)\end{bmatrix} \begin{bmatrix}\mathbf{r} \\ \boldsymbol{\Delta}\mathbf{b}\end{bmatrix} - \mathbf{f}_i \right\|_2^2 - \lambda \mathbf{g}_i^{\mathbf{T}} \boldsymbol{\Delta}\mathbf{b} \\ &= \|\mathbf{Z}\mathbf{m} - \mathbf{f}_i\|_2^2 - \lambda \mathbf{U}^{\mathbf{T}}\mathbf{m},\end{aligned} \tag{A-7}$$

where

$$\mathbf{Z} = \begin{bmatrix}\mathbf{H}(\mathbf{b}_0) & -\mathbf{W}(\mathbf{b}_0)\end{bmatrix}, \mathbf{U} = \begin{bmatrix}\mathbf{0} \\ \mathbf{g}_i\end{bmatrix}, \text{ and } \mathbf{m} = \begin{bmatrix}\mathbf{r} \\ \boldsymbol{\Delta}\mathbf{b}\end{bmatrix}. \tag{A-8}$$

Denoting the result as $\Phi(\mathbf{m})$ and expanding the terms give

$$\begin{aligned}\Phi(\mathbf{m}) &= \left(\mathbf{m}^{\mathbf{T}}\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\mathbf{m} - 2\mathbf{m}^{\mathbf{T}}\mathbf{Z}^{\mathbf{T}}\mathbf{f}_i + ||\mathbf{f}_i||_2^2\right) - \lambda \mathbf{m}^{\mathbf{T}}\mathbf{U} \\ &= \mathbf{m}^{\mathbf{T}}\left(\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\mathbf{m} - 2\mathbf{Z}^{\mathbf{T}}\mathbf{f}_i - \lambda\mathbf{U}\right) + ||\mathbf{f}_i||_2^2.\end{aligned} \tag{A-9}$$

Neglect the constant term $||\mathbf{f}_i||_2^2$, our goal is to solve the quadratic optimization problem

$$\Phi(\mathbf{m}) = \mathbf{m}^{\mathbf{T}}\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\mathbf{m} - 2\mathbf{m}^{\mathbf{T}}\mathbf{Z}^{\mathbf{T}}\mathbf{f}_i - \lambda \mathbf{m}^{\mathbf{T}}\mathbf{U}, \tag{A-10}$$

whose Hessian is given by $\mathbf{Z}^{\mathbf{T}}\mathbf{Z}$ and is thus positive semidefinite. A solution to this problem is given by any solution of the linear system

$$\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\mathbf{m} = \mathbf{Z}^{\mathbf{T}}\mathbf{f}_i + \frac{\lambda}{2}\mathbf{U}. \tag{A-11}$$

APPENDIX B

We finally present a formal justification of the approximation $\mathbf{H}(\mathbf{b}_0+\mathbf{\Delta b})\mathbf{r} \approx \mathbf{H}(\mathbf{b}_0)\mathbf{r}$, when we assume that $||\mathbf{\Delta b}||_2 << ||\mathbf{b}_0||_2$, which was used to obtain the objective function (18). In fact, to be precise, we will also assume that $||\mathbf{r}||_2 << ||\mathbf{b}_0||_2$, but this is a standard assumption in linearized waveform inversion whose validity we will not try to address here.

We can express the Hessian in terms of the Born modeling and the adjoint Born modeling operators as

$$\mathbf{H}(\mathbf{b}_0 + \mathbf{\Delta b}) = \mathbf{L}(\mathbf{b}_0 + \mathbf{\Delta b})^{\mathbf{T}}\mathbf{L}(\mathbf{b}_0 + \mathbf{\Delta b}). \tag{A-12}$$

Next, we can expand the Born modeling operator around $\mathbf{b}_0$ and apply it to the reflectivity $\mathbf{r}$ to obtain

$$\begin{aligned}\mathbf{L}(\mathbf{b}_0 + \mathbf{\Delta b})\mathbf{r} &= \mathbf{L}(\mathbf{b}_0)\mathbf{r} + \frac{\partial \mathbf{L}}{\partial \mathbf{b}}(\mathbf{b}_0; \mathbf{\Delta b}, \mathbf{r}) + O(||\mathbf{\Delta b}||^2, \mathbf{r}) \\ &= \mathbf{L}(\mathbf{b}_0)\mathbf{r} + \mathbf{T}(\mathbf{b}_0, \mathbf{r})\mathbf{\Delta b} + O(||\mathbf{\Delta b}||^2, \mathbf{r}),\end{aligned} \tag{A-13}$$

where $\mathbf{T}$ represents the tomographic operator introduced before, which is the derivative of the Born modeling operator with respect to the background model. One should note that $\mathbf{T}$ is a bilinear operator with respect to both $\mathbf{\Delta b}$ and $\mathbf{r}$.

We can obtain the corresponding approximation for the Hessian by pre-multiplying by the adjoint Born operator:

$$\mathbf{H}(\mathbf{b}_0 + \mathbf{\Delta b})\mathbf{r} = \mathbf{L}^{\mathbf{T}}(\mathbf{b}_0)\mathbf{L}(\mathbf{b}_0)\mathbf{r} + 2\mathbf{L}^{\mathbf{T}}(\mathbf{b}_0)\mathbf{T}(\mathbf{b}_0, \mathbf{r})\mathbf{\Delta b} + O(||\mathbf{\Delta b}||^2, \mathbf{r}). \tag{A-14}$$

which first justifies the approximation

$$\mathbf{H}(\mathbf{b}_0 + \mathbf{\Delta b})\mathbf{r} \approx \mathbf{H}(\mathbf{b}_0)\mathbf{r} + 2\mathbf{L}^{\mathbf{T}}(\mathbf{b}_0)\mathbf{T}(\mathbf{b}_0, \mathbf{r})\mathbf{\Delta b} \tag{A-15}$$

Next, since the second term in the right-hand side of equation (A-15) is bilinear with respect to both $\mathbf{\Delta b}$ and $\mathbf{r}$, and because everything is finite dimensional in the discretized setting, we know that there exists a linear operator $\tilde{\mathbf{T}}(\mathbf{b}_0)$ such that

$$\mathbf{L}^{\mathbf{T}}(\mathbf{b}_0)\mathbf{T}(\mathbf{b}_0, \mathbf{r})\mathbf{\Delta b} = (\tilde{\mathbf{T}}(\mathbf{b}_0)\mathbf{\Delta b})\mathbf{r}, \tag{A-16}$$

where $\tilde{\mathbf{T}}(\mathbf{b}_0)\mathbf{\Delta b}$ is a matrix, i.e. $\tilde{\mathbf{T}}(\mathbf{b}_0)$ should be thought of as a rank 3 tensor (or an abstract linear operator that maps vectors $\mathbf{\Delta b}$ to matrices). In particular this implies a bound of the form

$$||\mathbf{L}^{\mathbf{T}}(\mathbf{b}_0)\mathbf{T}(\mathbf{b}_0, \mathbf{r})\mathbf{\Delta b}||_2 \leq C||\mathbf{\Delta b}||_2||\mathbf{r}||_2, \tag{A-17}$$

where C is some constant that depends only on $\mathbf{b}_0$. This immediately gives rise to the easy lemma:

Lemma 1. Let $||\mathbf{H}(\mathbf{b}_0)\mathbf{r}||_2 = \delta$. Then for every $\mathbf{r}$ and $\gamma > 0$, there exists an $\epsilon > 0$, such that if $||\mathbf{\Delta b}||_2 < \epsilon$, $||\mathbf{L^T}(\mathbf{b}_0)\mathbf{T}(\mathbf{b}_0, \mathbf{r})\mathbf{\Delta b}||_2 < \gamma\delta$.

In fact if $\mathbf{r}$ is bounded we also get the slightly stronger result:

Lemma 2. Let $||\mathbf{H}(\mathbf{b}_0)\mathbf{r}||_2 = \delta$, and suppose $\mathbf{r}$ is bounded. Then for every $\gamma > 0$, there exists an $\epsilon > 0$, such that if $||\mathbf{\Delta b}||_2 < \epsilon$, $||\mathbf{L^T}(\mathbf{b}_0)\mathbf{T}(\mathbf{b}_0, \mathbf{r})\mathbf{\Delta b}||_2 < \gamma\delta$, $\forall\, \mathbf{r}$ that satisfies the boundedness assumption.

The proofs of the above lemmas are obvious and hence skipped. The main point is that one indeed has a regime given by $||\mathbf{\Delta b}||_2 < \epsilon$, where the desired approximation will hold, the accuracy being controlled by the parameter ϵ.

REFERENCES

Almomin, A., 2013, Accurate implementation of two-way wave-equation operators.

Barnier, G., and A. Almomin, 2014, Tutorial on two-way wave equation operators for acoustic, isotropic, constant-density media.

Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

Cabrales-Vargas, A., 2018a, Extended linearized waveform inversion with velocity updating.

———, 2018b, Improving reflectivity using linearized waveform inversion with velocity updating.

Dai, W., Y. Huang, and G. Schuster, 2013, Least-squares reverse time migration of marine data with frequency-selection encoding: SEG Technical Program Expanded Abstracts, 3231–3236.

Duquet, B., K. Marfurt, and J. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity and subsurface illumination: Geophysics, **65**, 1195–1209.

Fletcher, R., D. Nichols, R. Bloor, and R. Coates, 2016, Least-squares migation - data domain versus image domain using point spread functions: The Leading Edge, **35**, no. 2, 157–162.

Gazdag, J., and E. Carrizo, 1986, On reverse-time migration: Geophysical Prospecting, **34**, 822–832.

Kosloff, D., and E. Baysal, 1983, Migration with the full acoustic wave equation: Geophysics, **48**, 677–687.

Luo, S., and D. Hale, 2014, Least-squares migration in the presence of velocity errors: Geophysics, **79**, no. 6, s153–s161.

Nemeth, T., and G. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, no. 1, 208–221.

Nocedal, J., and S. Wright, 2006, Numerical optimization: Springer.

Schuster, G., 2017, Seismic inversion: Society of Exploration Geophysicists.

Wong, M., B. Biondi, and S. Ronen, 2015, Imaging with primaries and free-surface multiples by joint least-squares reverse time migration: Geophysics, **80**, no. 6, S223–S235.

Yang, W., Y. Li, A. Cheng, Y. Liu, and L. Dong, 2018, Least-squares reverse time

migration with velocity errors: SEG Technical Program Expanded Abstracts, 4256–4260.