Linearized Waveform Inversion with Velocity Updating: Theory and first results

Alejandro Cabrales-Vargas, Biondo Biondi, and Robert Clapp

ABSTRACT

Linearized Waveform Inversion or least-squares migration is a process that aims at obtaining a better estimation of the subsurface reflectivity, in comparison with conventional migration. During the process, the background model (velocity or slowness) remains invariant. Only the reflectivity component is updated.
In this report we revisit the Linearized Waveform Inversion with Velocity Updating theory introduced in a previous report, and present the first synthetic examples. The method introduces controlled updates to the background model during the reflectivity inversion, correcting for slowness inaccuracies that negatively affect seismic amplitudes during conventional linearized waveform inversion. The method incorporates Wave-Equation Migration Velocity Analysis to transform such background model updates into perturbations in the image.

INTRODUCTION

Imaging complex oil & gas reservoirs with seismic methods demands solutions beyond kinematically accurate subsurface images. Beyond identification of reservoir traps, proper characterization of oil-bearing rock facies is vital for optimal drilling programs and exploitation. One important tool in the achievement of these goals is the variation of seismic amplitudes as a function of the rock and fluid type.

For several years, the interpretation of amplitude variations has relied on "true-amplitude" Kirchhoff migration images. The high-frequency approximation assumption entailed by Kirchhoff modeling and migration algorithms allows splitting the solution into two components: a kinematic component, and an amplitude component. The former is obtained by means of traveltime computations, resulting in an image with correctly positioned seismic events in relatively simple geology. The latter (amplitude coefficients) is usually an approximation to the transport equation solution.

Kirchhoff-based solutions present an important caveat: the high-frequency approximation fails in the presence of strong velocity contrasts and complex geology. Downward continuation methods and two-way wave equation methods are thereby mandatory for adequately addressing such complexities. Nonetheless, the final image usually is only kinematically correct. One reason is that the imaging condition constitutes a zero-lag crosscorrelation of the source wavefield and the receiver wavefield

(Claerbout, 1971). The amplitudes can be balanced by implementing deconvolution imaging conditions (e.g. Guitton et al., 2007), but the amplitude still remains inaccurate.

One important limitation of the migration methods, regardless of the amplitude treatment, is that they merely constitute the first step of the inversion process aimed at estimating the subsurface reflectivity:

$$\tilde{\mathbf{m}} = \mathbf{L}'\mathbf{d}_0, \tag{1}$$

where $\mathbf{d}_0$ represents the seismic data collected at discrete surface locations, $\tilde{\mathbf{m}}$ represents here the estimated reflectivity, and $\mathbf{L}'$ represents the *adjoint* of the Born modeling operator, better known as *migration* (Claerbout, 1992). We denominate $\tilde{\mathbf{m}}$ as the *migrated image*. Born modeling constitutes a linearization of the full modeling operator, $\mathcal{L}$, around a background velocity (or slowness) field, with the reflectivity acting as a perturbation in such background field.

Unfortunately, the amplitudes of the migration image are often not representative of the geologic variations. The problem is further exacerbated by irregular and/or sparse acquisition geometries (frequently due to obstacles and/or limited budget) and limited frequency bandwidth of the data. As a consequence, the resulting image often constitutes a blurred version of the subsurface reflectivity: the imperfect acquisition geometry and limited frequency content negatively impact both the seismic resolution and amplitude preservation.

Linearized waveform inversion (LWI) (*a.k.a.* least-squares migration) (e.g. Nemeth and Schuster, 1999; Ronen and Liner, 2000; Clapp, 2005; Valenciano, 2008; Tang, 2011b; Dai et al., 2010; Fletcher et al., 2016) constitutes an iterative process for approximating the inverse of seismic modeling, hence recovering a better reflectivity estimation than conventional migration. It can be expressed as the solution of the normal equations, formally cast as

$$\mathbf{L}'\mathbf{L}\mathbf{m} = \mathbf{L}'\mathbf{d}_0, \tag{2}$$

where $\mathbf{m}$ constitutes the subsurface reflectivity. Solving Equations 2 is equivalent to minimizing the following objective function in data space

$$\Phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{L}\mathbf{m} - \mathbf{d}_0\|_2^2, \tag{3}$$

or in model space

$$\Phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{H}\mathbf{m} - \tilde{\mathbf{m}}\|_2^2, \tag{4}$$

where $\mathbf{d} = \mathbf{L}\mathbf{m}$ constitutes synthetic data, and $\mathbf{H}$ is the *Hessian* operator associated to the misfit function Φ (*a.k.a. Gauss Newton Hessian*). Note that in Equation 3 we aim at fitting synthetic data to the recorded data, whereas in Equation 4 we aim at fitting a reflectivity model to the migrated image. Both constitute different versions of the same optimization problem, which is *linear* because the Born operator and the Hessian are linear with respect to the reflectivity model. In other words, such operators are *independent* of the reflectivity model.

LWI has been so far parameterized exclusively in terms of the subsurface reflectivity. The velocity field is assumed to be either correct or inaccurate to some degree (e.g. Luo and Hale, 2014), but either way fixed during the optimization process. It does not participate in the reflectivity estimation during LWI. Thus, the obvious question is, can we push the LWI capabilities further by allowing velocity (or slowness) to participate in the optimization process?

We developed Linearized waveform inversion with velocity updating (LWIVU) to incorporate a perturbation component of the velocity field into the optimization process (Cabrales-Vargas et al., 2016). The objective is to obtain better estimations of the reflectivity with respect to conventional LWI. The updates in the velocity (or slowness) field are intended to be perturbations small enough to maintain the linearity of the inversion. Moreover, such perturbations are not incorporated back to the background velocity field, but merely used to compute perturbations in the image that would increase the accuracy of the inverted reflectivity. With this approach, we exploit the fact that, in real data, the "correct velocity field" is accurate enough for positioning seismic events, but remaining inaccuracies can affect the amplitudes.

LWIVU can be derived from a simplification of the nonlinear optimization process known as *full-waveform inversion* (FWI), by splitting the Hessian into Gauss-Newton Hessian and *wave-equation migration velocity analysis* (WEMVA) Hessian, and adding a WEMVA-based constraint that maximizes the stacking power. In the next section we revisit the original derivation introduced in our previous report (Cabrales-Vargas et al., 2016). Next, we present synthetic examples in a two-layer model with a Gaussian slowness anomaly. Next, we analyse the computational costs of LWIVU compared to conventional LWI. Finally, we present the conclusions of this report.

THEORY

Derivation of LWIVU from FWI

Full-waveform inversion (FWI) (Tarantola, 1984; Virieux and Operto, 2009; Fichtner, 2011; Biondi and Almomin, 2014) is a nonlinear optimization scheme that minimizes the misfit between the recorded seismic data, $\mathbf{d}_r$, and modeled data, $\mathbf{d}$, with respect to a subsurface parameters model, $\mathbf{m}$,

$$\Phi_{FWI}(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - \mathbf{d}_r\|_2^2. \tag{5}$$

where $\mathbf{d} = \mathcal{L}(\mathbf{m})$. As aforementioned, $\mathcal{L}$ constitutes the *non-linear seismic modeling operator*. It is worth remarking that the non-linearity of this operator is with respect to the model parameters, $\mathbf{m}$. On the contrary, the operator is linear with respect to the source term (Almomin, 2013).

We can apply the Gauss-Newton method to minimizing the misfit function in Equation 5. Firstly, notice that a necessary condition is that the gradient of the misfit function vanishes at a minimum (Fichtner, 2011),

$$\nabla\Phi_{FWI}(\mathbf{m}_{min}) = 0, \tag{6}$$

where $\mathbf{m}_{min}$ constitutes such a minumum. Now we can expand the gradient in Equation 6 in Taylor series around a nearby model, $\mathbf{m}_i$, such that the difference $\delta\mathbf{m} = \mathbf{m}_{min} - \mathbf{m}_i$ is small enough to allowing us to truncate the series at the second term:

$$\nabla\Phi_{FWI}(\mathbf{m}_{min}) \approx \nabla\Phi_{FWI}(\mathbf{m}_i) + \mathbf{H}(\mathbf{m}_i)\delta\mathbf{m} = 0, \tag{7}$$

where $\mathbf{H}$ represents the FWI Hessian operator. It is the second derivative of the misfit function with respect to the model parameters. We use this result to estimate the model update, $\delta\mathbf{m}$, by solving the linear system

$$\mathbf{H}(\mathbf{m}_i)\delta\mathbf{m} = -\nabla\Phi_{FWI}(\mathbf{m}_i), \tag{8}$$

which is known as the *Newton's equation.* In the FWI problem we solve Equation 8 as part of an iterative non-linear optimization scheme aimed at updating the model parameters using $\mathbf{m}_{i+1} = \mathbf{m}_i + \delta\mathbf{m}$, repeating the process until desired convergence is reached.

For the LWIVU derivation, we adopt the Barnier and Almomin (2014) notation, separating the model parameters in a low-wavenumber component, $\mathbf{b}$, and a high-wavenumber component, $\mathbf{r}$,

$$\mathbf{m} = \mathbf{b} + \mathbf{r}. \tag{9}$$

We refer to $\mathbf{b}$ and $\mathbf{r}$ as the *background* and *reflectivity* components, respectively. Now we perturb each component,

$$\mathbf{m} = \mathbf{b}_0 + \boldsymbol{\Delta}\mathbf{b} + \mathbf{r}_0 + \boldsymbol{\Delta}\mathbf{r}, \tag{10}$$

where $\mathbf{b}_0$ constitutes the *most background* model, $\boldsymbol{\Delta}\mathbf{b}$ is the *perturbation in the background* model, $\mathbf{r}_0$ is the *background reflectivity*, and $\boldsymbol{\Delta}\mathbf{r}$ is the *perturbation in the reflectivity.* In practice, we assume that the model parameters (*e.g.*, slowness) have transitions smooth enough to allowing us to neglect the background reflectivity, hence $\mathbf{r}_0 = \mathbf{0}$. Therefore, the reflectivity image is defined exclusively by $\boldsymbol{\Delta}\mathbf{r}$. For such reason we sometimes refer to it as "the image".

With these considerations we can recast the Newton's equation in such a way that we keep $\mathbf{b}_0$ unchanged during the inversion, and only update the perturbations by setting $\delta\mathbf{m} = \boldsymbol{\Delta}\mathbf{b} + \boldsymbol{\Delta}\mathbf{r}$. Thus, Equation 8 becomes

$$\mathbf{H}(\mathbf{b}_0)[\boldsymbol{\Delta}\mathbf{b} + \boldsymbol{\Delta}\mathbf{r}] = -\nabla\Phi_{FWI}(\mathbf{b}_0). \tag{11}$$

We first interpret the meaning of the right-hand side of Equation 11. Back to Equation 5, let us set $\mathbf{m} = \mathbf{b}_0$,

$$\Phi_{FWI}(\mathbf{b}_0) = \frac{1}{2}\|\mathcal{L}(\mathbf{b}_0) - \mathbf{d}_r\|_2^2, \tag{12}$$

with the corresponding gradient becoming

$$\nabla\Phi_{FWI}(\mathbf{b}_0) = \left[\frac{\partial\mathcal{L}(\mathbf{m})}{\partial\mathbf{m}}\bigg|_{\mathbf{m}=\mathbf{b}_0}\right]'\big[\mathcal{L}(\mathbf{b}_0) - \mathbf{d}_r\big], \tag{13}$$

where $'$ represents the adjoint. The synthetic data evaluated at the most background model contains transmitted and direct waves, *but no reflections*, while the recorded data contains all the events. Therefore, this data difference corresponds to the negative of the recorded data after removing the transmitted and direct waves, *i.e.*, the negative of the reflection data, $\mathbf{d}_0 = \mathbf{d}_r - \mathcal{L}(\mathbf{b}_0)$. The derivative of the synthetic data evaluated at the most background model is the *Born modeling approximation.* The adjoint of this operator is *reverse-time migration* (RTM) (Barnier and Almomin, 2014). Therefore, the gradient in Equation 13 represents the negative of the reverse-time migration image,

$$\nabla\Phi(\mathbf{b}_0) = -\mathbf{\Delta r}_{mig}, \tag{14}$$

where we represent the migrated image as $\mathbf{\Delta r}_{mig}$ to be consistent with the notation employed throughout this report. The migrated image (and hence the gradient) remains unchanged during the optimization process.

Now we interpret the left-hand side of Equation 11. According to Biondi et al. (2015) we can split the full Hessian $\mathbf{H}$ into the so-called Gauss-Newton Hessian, $\mathbf{H}_{GN}$, and the wave-equation migration velocity analysis (WEMVA) Hessian, $\mathbf{H}_W$:

$$\mathbf{H} = \mathbf{H}_{GN} + \mathbf{H}_W. \tag{15}$$

Substituting Equation 14 and Equation 15 into Equation 11 we obtain

$$(\mathbf{H}_{GN} + \mathbf{H}_W)(\mathbf{\Delta b} + \mathbf{\Delta r}) = \mathbf{\Delta r}_{mig}. \tag{16}$$

We can explicitely apply each Hessian to the perturbations, regrouping terms in a convenient manner:

$$(\mathbf{H}_{GN}\mathbf{\Delta r} + \mathbf{H}_W\mathbf{\Delta b}) + (\mathbf{H}_{GN}\mathbf{\Delta b} + \mathbf{H}_W\mathbf{\Delta r}) = \mathbf{\Delta r}_{mig}. \tag{17}$$

We first analyze the first part of the left-hand side of Equation 17. The first term, $\mathbf{H}_{GN}\mathbf{\Delta r}$, is the Gauss-Newton Hessian applied to the perturbation of the reflectivity. The Gauss-Newton Hessian constitutes the adjoint of the Born modeling operator followed by the corresponding forward operator: $\mathbf{H}_{GN} = \mathbf{L}^T\mathbf{L}$. Fitting this term alone to the migrated image represents conventional LWI in *model space* (Valenciano et al., 2009), to be distinguished from the LWI formulation in *data space* (Nemeth and Schuster, 1999) (both methods are sketched out in Equations 1 to 3). Therefore, this term yields a reflectivity estimation that is more accurate than the conventionally migrated image, in terms of amplitude and seismic resolution. The second term, $\mathbf{H}_W\mathbf{\Delta b}$, is the WEMVA operator applied to the perturbation of the background model, which yields a perturbation of the reflectivity image. Such term is what we need to implement the proposed method.

Now we analyze the second part of the left-hand side of Equation 17. The first term, $\mathbf{H}_{GN}\boldsymbol{\Delta}\mathbf{b}$, is the Gauss-Newton Hessian applied to the perturbation of the background model. Insofar as our objective is to account for reflectivity estimations, we can neglect this term because improving the background model is not part of this objective. The second term, $\mathbf{H}_W\boldsymbol{\Delta}\mathbf{r}$, is the WEMVA operator applied to the perturbation of the reflectivity. This term accounts for second-order scattering effects, such as multiples. We do not consider such effects for LWIVU, therefore we can neglect this term too.

After these considerations, we simplify Equation 17 to obtain

$$\mathbf{H}_{GN}\boldsymbol{\Delta}\mathbf{r} + \mathbf{H}_W\boldsymbol{\Delta}\mathbf{b} = \begin{bmatrix}\mathbf{H}_{GN} & \mathbf{H}_W\end{bmatrix}\begin{bmatrix}\boldsymbol{\Delta}\mathbf{r}\\ \boldsymbol{\Delta}\mathbf{b}\end{bmatrix} \approx \boldsymbol{\Delta}\mathbf{r}_{mig}. \tag{18}$$

From Equation 18 we can cast an optimization problem with the following misfit function,

$$\Phi_1(\boldsymbol{\Delta}\mathbf{r}, \boldsymbol{\Delta}\mathbf{b}; \mathbf{b}_0) = \frac{1}{2}\|\mathbf{H}_{GN}\boldsymbol{\Delta}\mathbf{r} + \mathbf{H}_W\boldsymbol{\Delta}\mathbf{b} - \boldsymbol{\Delta}\mathbf{r}_{mig}\|_2^2, \tag{19}$$

where the optimization parameters are $\boldsymbol{\Delta}\mathbf{r}$ and $\boldsymbol{\Delta}\mathbf{b}$, for a seismic experiment realization with the most background model $\mathbf{b}_0$. Both Hessians are independent of the optimization parameters; they only depend on $\mathbf{b}_0$. Hence, the optimization problem portrayed by Equation 19 is *linear*. We have labeled this misfit function as Φ_1 because an additional constraint is required to update $\boldsymbol{\Delta}\mathbf{b}$ in synergy with the original purpose of producing more focused images. One way to ensure this happens is enforcing the maximization of stacking power, or conversely, the minimization of the *negative* of stacking power. On these grounds we include an additional term in the misfit function,

$$\Phi_2(\boldsymbol{\Delta}\mathbf{b}; \mathbf{b}_0) = -\frac{\lambda^2}{2}\|\mathbf{H}_W\boldsymbol{\Delta}\mathbf{b} + \boldsymbol{\Delta}\mathbf{r}_{mig}\|_2^2, \tag{20}$$

where λ is a parameter that allows us to control the level of stacking power maximization. In order to analyze the interaction between the misfit functions Φ_1 and Φ_2, we express Equation 19 as an optimization problem aimed at fitting the reflectivity model to an improved migrated image $\widehat{\boldsymbol{\Delta}\mathbf{r}}_{mig}(\boldsymbol{\Delta}\mathbf{b})$, such that

$$\widehat{\boldsymbol{\Delta}\mathbf{r}}_{mig}(\boldsymbol{\Delta}\mathbf{b}) = \boldsymbol{\Delta}\mathbf{r}_{mig} - \mathbf{H}_W\boldsymbol{\Delta}\mathbf{b}, \tag{21}$$

therefore Equation 19 becomes

$$\Phi_1(\boldsymbol{\Delta}\mathbf{r}, \boldsymbol{\Delta}\mathbf{b}; \mathbf{b}_0) = \frac{1}{2}\|\mathbf{H}_{GN}\boldsymbol{\Delta}\mathbf{r} - \widehat{\boldsymbol{\Delta}\mathbf{r}}_{mig}(\boldsymbol{\Delta}\mathbf{b})\|_2^2, \tag{22}$$

According to this definition, once $\boldsymbol{\Delta}\mathbf{b}$ satisfies the misfit function, Φ_2 (thus maximizing the stacking power), it should yield an improved migrated image, $\widehat{\boldsymbol{\Delta}\mathbf{r}}_{mig}$. However, the maximizing effect gained by $\boldsymbol{\Delta}\mathbf{b}$ precisely *removes* such correction from such migrated image, because of the *minus* sign in Equation 21. Such a scheme would harm the reflectivity estimation. Hence, it becomes logical to re-define the improved migrated image as the term to be maximized inside the norm in Equation 20,

$$\widehat{\boldsymbol{\Delta}\mathbf{r}}_{mig}(\boldsymbol{\Delta}\mathbf{b}) = \boldsymbol{\Delta}\mathbf{r}_{mig} + \mathbf{H}_W\boldsymbol{\Delta}\mathbf{b}, \tag{23}$$

thus the LWIVU misfit function becomes

$$\begin{aligned}\Phi(\mathbf{\Delta r},\mathbf{\Delta b};\mathbf{b}_0) &= \Phi_1(\mathbf{\Delta r},\mathbf{\Delta b};\mathbf{b}_0)+\Phi_2(\mathbf{\Delta b};\mathbf{b}_0)\\ &= \frac{1}{2}\|\mathbf{H}_{GN}\mathbf{\Delta r}-\widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b})\|_2^2-\frac{\lambda^2}{2}\|\widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b})\|_2^2. \end{aligned}\tag{24}$$

Let us explicitely express Equation 24 in terms of $\mathbf{H\Delta b}$:

$$\Phi(\mathbf{\Delta r},\mathbf{\Delta b};\mathbf{b}_0)=\frac{1}{2}\|\mathbf{H}_{GN}\mathbf{\Delta r}-(\mathbf{H}_W\mathbf{\Delta b}+\mathbf{\Delta r}_{mig})\|_2^2-\frac{\lambda^2}{2}\|\mathbf{H}_W\mathbf{\Delta b}+\mathbf{\Delta r}_{mig}\|_2^2. \tag{25}$$

In order to implement gradient-based optimization schemes we obtain the gradient of the misfit function in Equation 25. Deriving with respect to the model parameters $\mathbf{\Delta r}$ and $\mathbf{\Delta b}$,

$$\nabla_{\mathbf{\Delta r}}\Phi=\mathbf{H}'_{GN}(\mathbf{H}_{GN}\mathbf{\Delta r}-\mathbf{H}_W\mathbf{\Delta b}-\mathbf{\Delta r}_{mig}), \tag{26}$$

$$\nabla_{\mathbf{\Delta b}}\Phi=\mathbf{H}'_{W}(\mathbf{H}_{GN}\mathbf{\Delta r}-\mathbf{H}_W\mathbf{\Delta b}-\mathbf{\Delta r}_{mig})-\lambda\mathbf{H}'_W\Big[\lambda(\mathbf{H}_W\mathbf{\Delta b}+\mathbf{\Delta r}_{mig})\Big], \tag{27}$$

the total gradient becomes

$$\nabla\Phi=\begin{bmatrix}\nabla_{\mathbf{\Delta r}}\Phi\\ \nabla_{\mathbf{\Delta b}}\Phi\end{bmatrix}=\begin{bmatrix}\mathbf{H}'_{GN} & \mathbf{0}\\ -\mathbf{H}'_W & -\lambda\mathbf{H}'_W\end{bmatrix}\begin{bmatrix}\mathbf{H}_{GN}\mathbf{\Delta r}-\mathbf{H}_W\mathbf{\Delta b}-\mathbf{\Delta r}_{mig}\\ \lambda(\mathbf{H}_W\mathbf{\Delta b}+\mathbf{\Delta r}_{mig}).\end{bmatrix} \tag{28}$$

The *adjoint* of LWIVU is the matrix-like operator at the right-hand side of Equation 28,

$$\begin{bmatrix}\mathbf{H}'_{GN} & \mathbf{0}\\ -\mathbf{H}'_W & -\lambda\mathbf{H}'_W\end{bmatrix}, \tag{29}$$

which is applied to the residuals vector. Note that such residuals correspond to the *fitting goals*, i.e., the optimization terms in Equation 25. The adjoint of the LWIVU adjoint operator constitutes the LWIVU *forward* operator:

$$\begin{bmatrix}\mathbf{H}_{GN} & -\mathbf{H}_W\\ \mathbf{0} & -\lambda\mathbf{H}_W\end{bmatrix}. \tag{30}$$

With the forward and adjoint operators we can set forth LWIVU as a gradient-based optimization scheme. We implement the optimization using the *conjugate directions method* (Claerbout, 2014). We discuss such implementation in the Appendix A. In the Appendix B we derive the WEMVA optimization problem using maximization of the stacking power, in an effort to provide the reader with additional tools for understanding the LWIVU derivation.

SYNTHETIC EXAMPLES: LWI VS. LWIVU

For the sake of illustration we implement LWIVU in a 2-D synthetic dataset obtained from a two-layer subsurface model with a Gaussian anomaly (Figure 1). The model space consists of 151 × 461 grid-points in depth (z) and distance (x), respectively, plus a 48-point haloing zone for the tapering boundary conditions. Grid spacing is $\Delta x = \Delta z = 20$ m. The slowness square of the upper and lower layers are 2.5×10^{-7} s^2/m^2 (2000 m/s) and 1.6×10^{-7} s^2/m^2 (2500 m/s), respectively. The Gaussian anomaly (Figure 2) has a minimum value of -2.3×10^{-8} s^2/m^2. It is added to the first layer. For this experimental setup, the initial slowness squared model consists of the two-layer model *without* the anomaly (Figure 3). This model corresponds to the most background model, $\mathbf{b}_0$, whereas the anomaly constitutes the perturbation of the background model, $\mathbf{\Delta b}$.

We obtained synthetic data by non-linear modeling using the true model and a constant model (the latter consists of the slowness of the upper layer only), and subtracted them to remove direct waves and only preserve reflections. The acquisition geometry consists of 31 sources spaced every 300 m with the first source at $x = 0$. Every source corresponds to 481 receivers spaced every 20 m. Source and receivers lie at the surface, $z = 0$.

We obtained the migrated image, $\mathbf{\Delta r}_{mig}$, by applying RTM to the data with the initial model (Figure 4). The Gauss-Newton Hessian and WEMVA Hessian are calculated with the most background model, $\mathbf{b}_0$. For the experiments shown in this report we only use the diagonal of the Gauss-Newton Hessian matrix, yielding a correction in the amplitude but not a resolution increase.

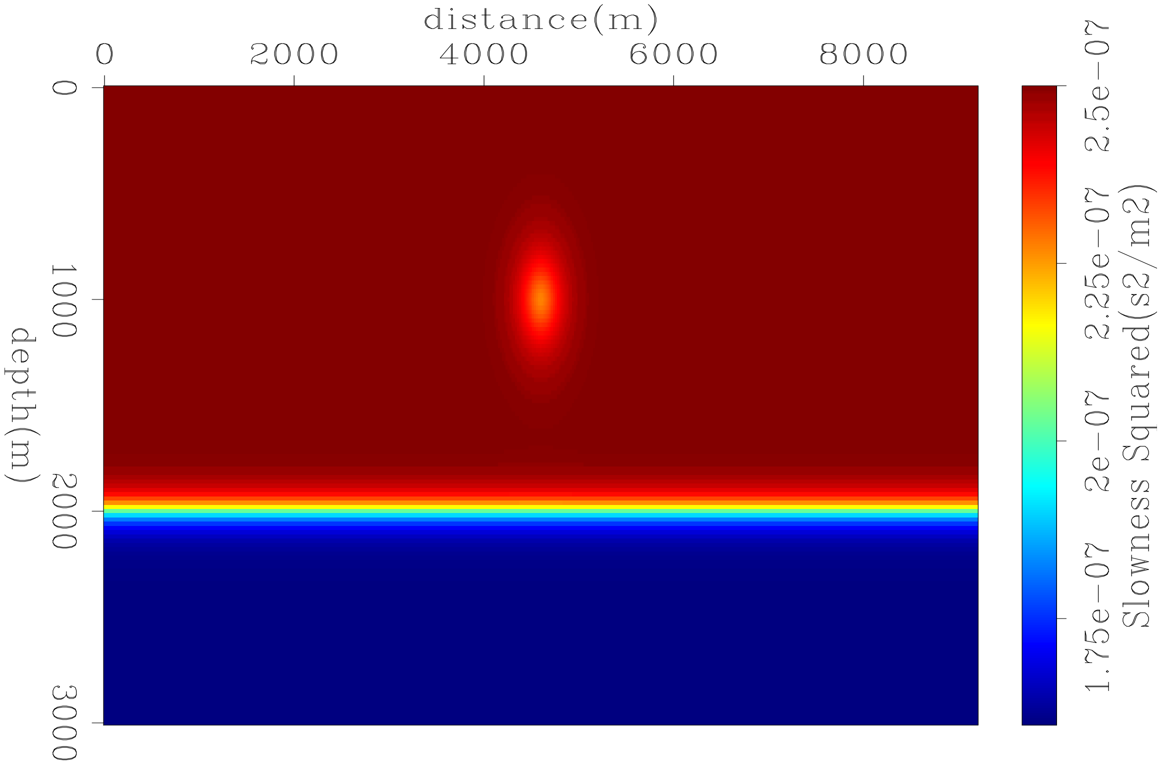

Figure 1: Slowness squared model with Gaussian anomaly. This model constitutes the real model [**ER**].

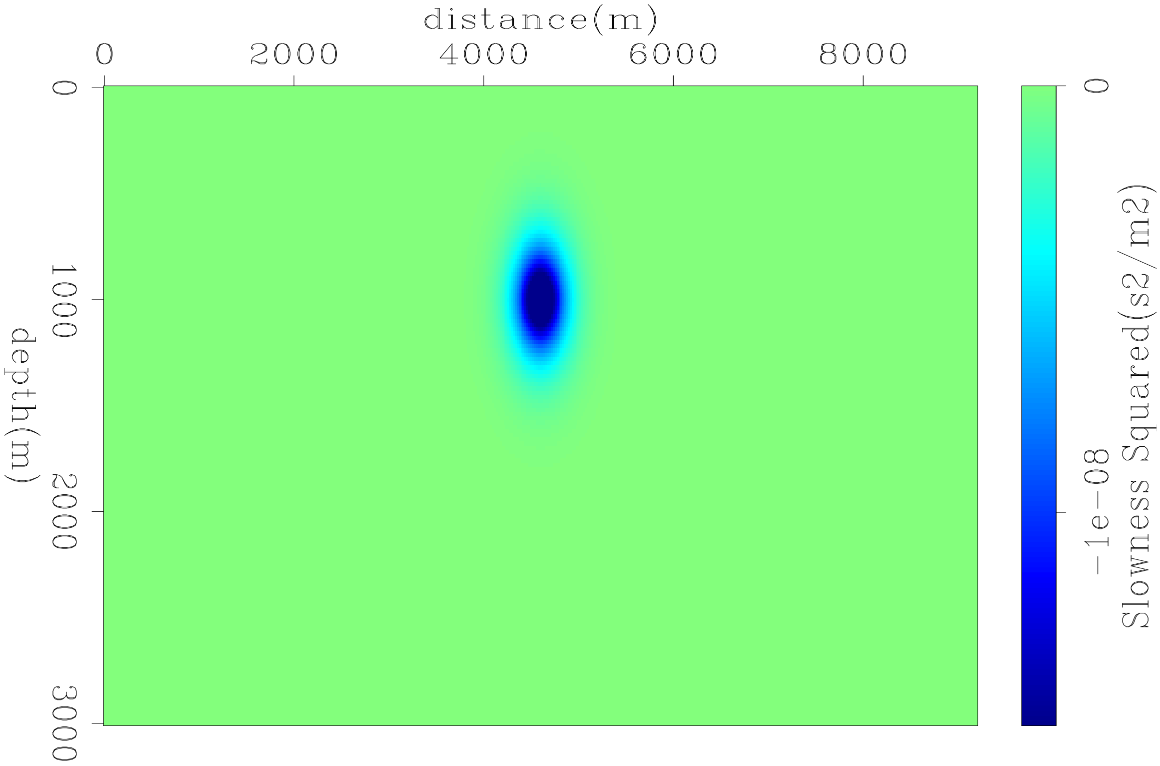

Figure 2: Slowness squared perturbation, $\mathbf{\Delta b}$: Gaussian anomaly [**ER**].

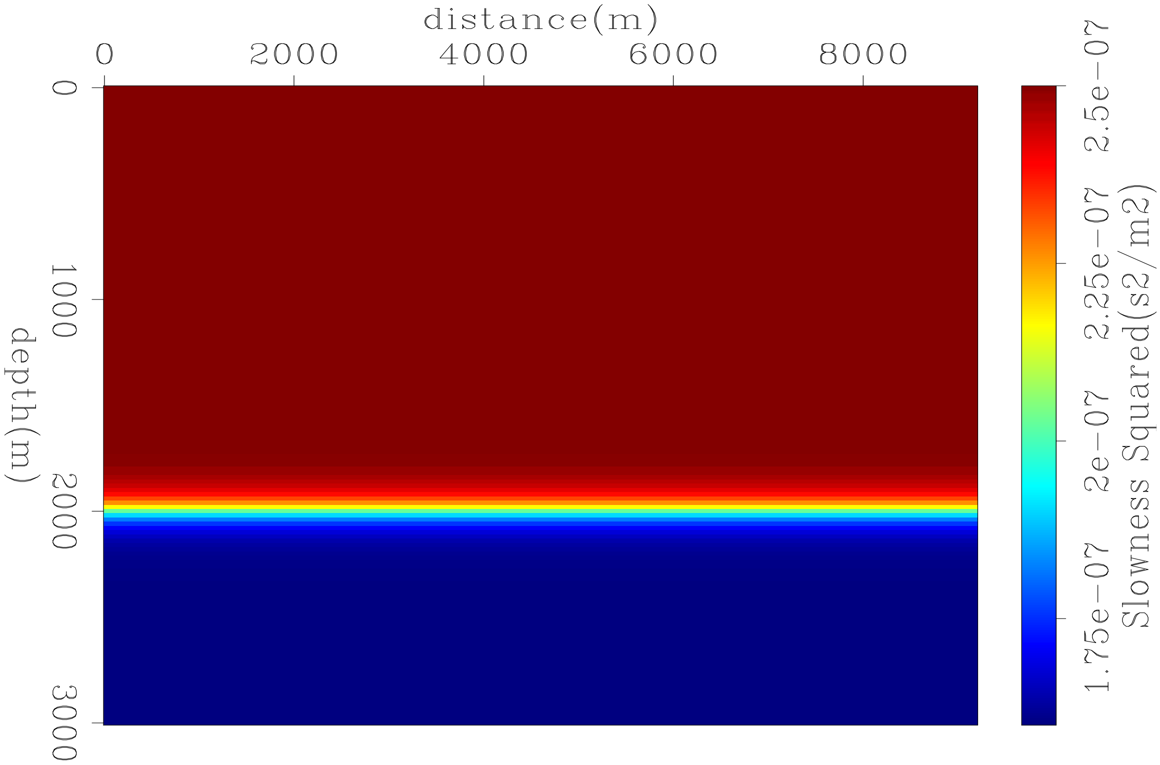

Figure 3: Slowness squared model without Gaussian anomaly. This model constitutes the most background model, $\mathbf{b}_0$ [**ER**].

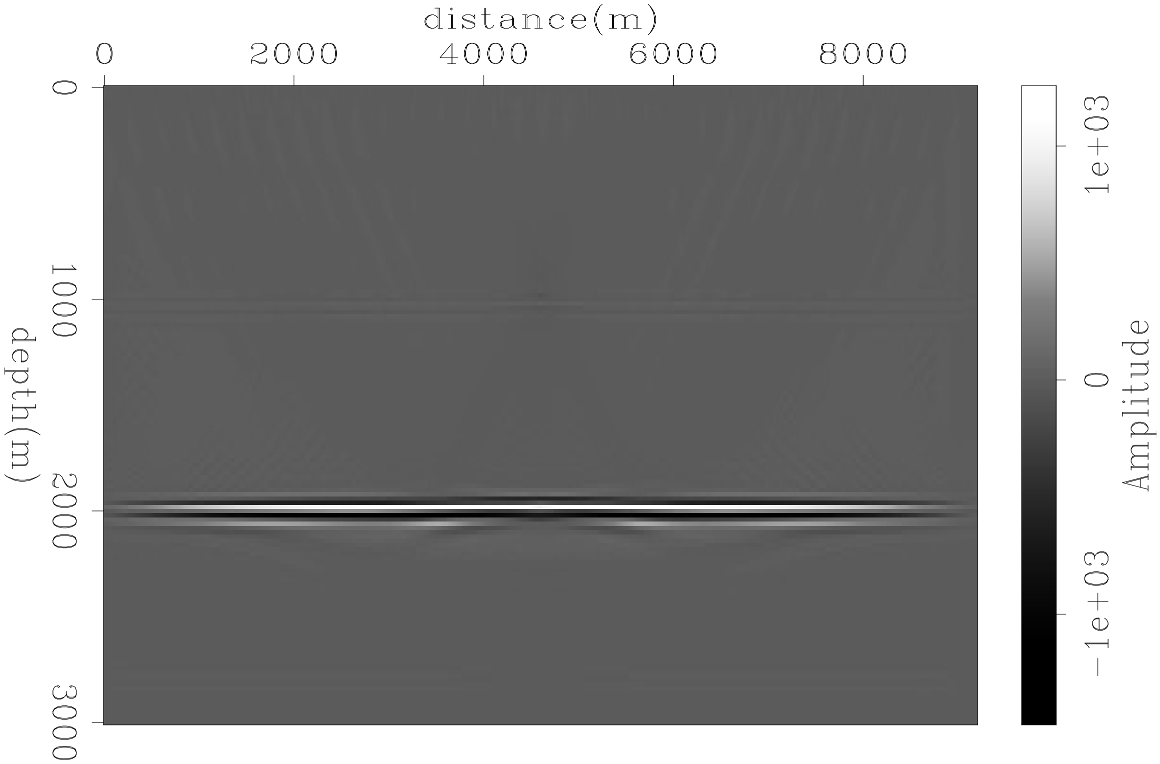

Figure 4: RTM image, $\mathbf{\Delta r}_{mig}$ [**CR**].

In the following plots we compare the LWI reflectivity (left panels) with the LWIVU reflectivity (right panels), and the true slowness squared perturbation (left panels) with the LWIVU perturbation (right panels), for different λ values. We obtained the LWI result in model space according to Equation 4, which becomes $\Phi(\mathbf{\Delta r}) = \frac{1}{2}\|\mathbf{H}_{GN}\mathbf{\Delta r} - \mathbf{\Delta r}_{mig}\|_2^2$ with the notation used in the theoretical section. For better appreciation we zoomed in on the reflectivity panels. We applied no clippling to the reflectivity panels. In the slowness squared perturbations panels we clipped to the minimum of the true anomaly. All the examples were run for 10 iterations, unless otherwise specified.

Figure 5 shows the reflectivity comparison for $\lambda = 0$, *i.e.*, without stacking power maximization. Note the amplitude gap at the center of the reflector in both panels caused by the incorrect slowness model. The excess in slowness square when ignoring the anomaly is too small to produce kinematic errors (*e.g.*, pull-ups), thus only amplitudes are affected. Two linear events cross at the center of the reflector gap. They are the smearing effect of conventional migration, and off-diagonal elements of the Gauss-Newton Hessian are needed to collapse such artifacts. LWIVU with $\lambda = 0$ did not recover the amplitude at the gap. Moreover, the perturbation in slowness squared is updated with the wrong polarity (positive) (Figure 6). The corresponding "improved" migrated image, $\widehat{\mathbf{\Delta r}}_{mig}$, has the gap reinforced, and the reflectivity is fitted to such undesired image.

Figure 7 shows the reflectivity comparison for $\lambda = 0.5$. The results are very similar to the previous case. However, the perturbation in slowness squared (Figure 8) is less positive than that of $\lambda = 0$ (Figure 6).

Figure 9 shows the reflectivity comparison for $\lambda = 1.0$. There is still no apparent change with respect to the previous cases, but the slowness squared perturbation flipped the sign and became negative (Figure 10). Imposing stacking power maximization pushes the perturbation of the image to fill the gap in the migrated image, which demands a negative $\mathbf{\Delta b}$. The reflectivity is fitted to this improved migrated image. Yet, 10 iterations are not enough to match the true slowness squared with $\lambda = 1.0$, then the improved migrated image does not have the gap filled, and as a consequence, neither does the reflectivity. Thus, we extended the inversion to 16 iterations. Note in Figure 11 that the gap previously observed in the right panel of Figure 9 has been healed. The shaded zone still observed at the reflector center can be the crossing linear events. The background in slowness squared in Figure 12 now satisfactorily resembles the true perturbation.

Figure 13 shows the reflectivity comparison for $\lambda = 1.5$. The time the gap has been healed, and the overall amplitude is increased. However, the perturbation appears to be slightly surpassing the true anomaly (Figure 14). Note the increase in amplitude of the subsamplig artifacts in the shallow part and the reflector leaking into the perturbation image. They appear not to be influencing the reflectivity. Using less iterations we can better fit the perturbation in slowness and reduce the artifacts, but then the gap is not completely healed.

Figure 15 shows the reflectivity comparison for $\lambda = 2.5$. The gap is eliminated, but the amplitude appears to be over-corrected (note the bias towards positive values in the scalebar). As expected, $\mathbf{\Delta b}$ surpasses the true slowness (Figure 16) further than $\lambda = 1.5$. Better results are obtained at the eighth iteration (Figure 17, note the unbiased scalebar), although $\mathbf{\Delta b}$ remains overestimated (Figure 18). Despite the over-correction, the overall amplitude decreased compared to $\lambda = 1.5$. The artifacts in $\mathbf{\Delta b}$ increased with respect to $\lambda = 1.5$ in both cases.

Finally, Figure 19 and Figure 20 are the comparisons of reflectivity and perturbation in slowness squared using $\lambda = 5.0$. The gap is over-corrected, whereas the amplitude significantly diminishes further away from the reflector center. In addition, the perturbation in slowness squared has been laterally compressed. The positive surrounding halo, the subsampling artifacts, and the leaked reflector are exacerbated compared to lower λ values. The reflectivity amplitudes can be impacted by such effects. The energy imaging condition (Rocha et al., 2016) can potentially attenuate the leaked reflectivity, whereas the subsampling artifacts can be attenuated by addressing the model null space.

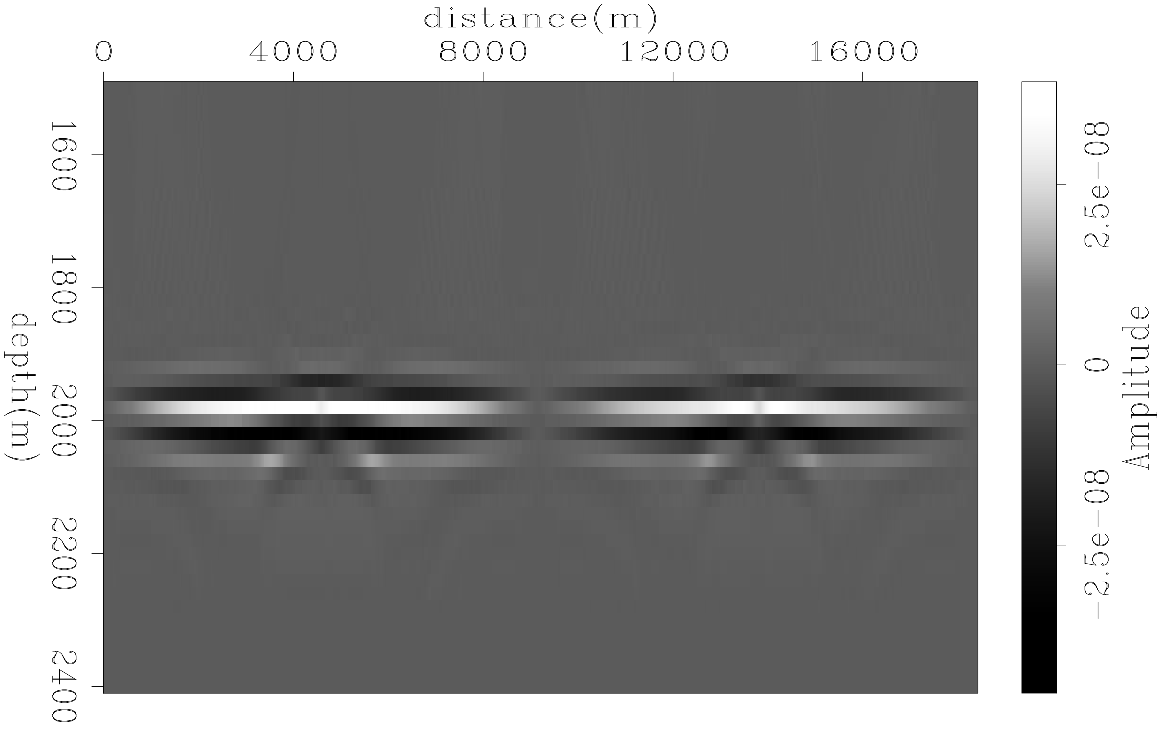

Figure 5: Reflectivity comparison, $\lambda = 0$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**].

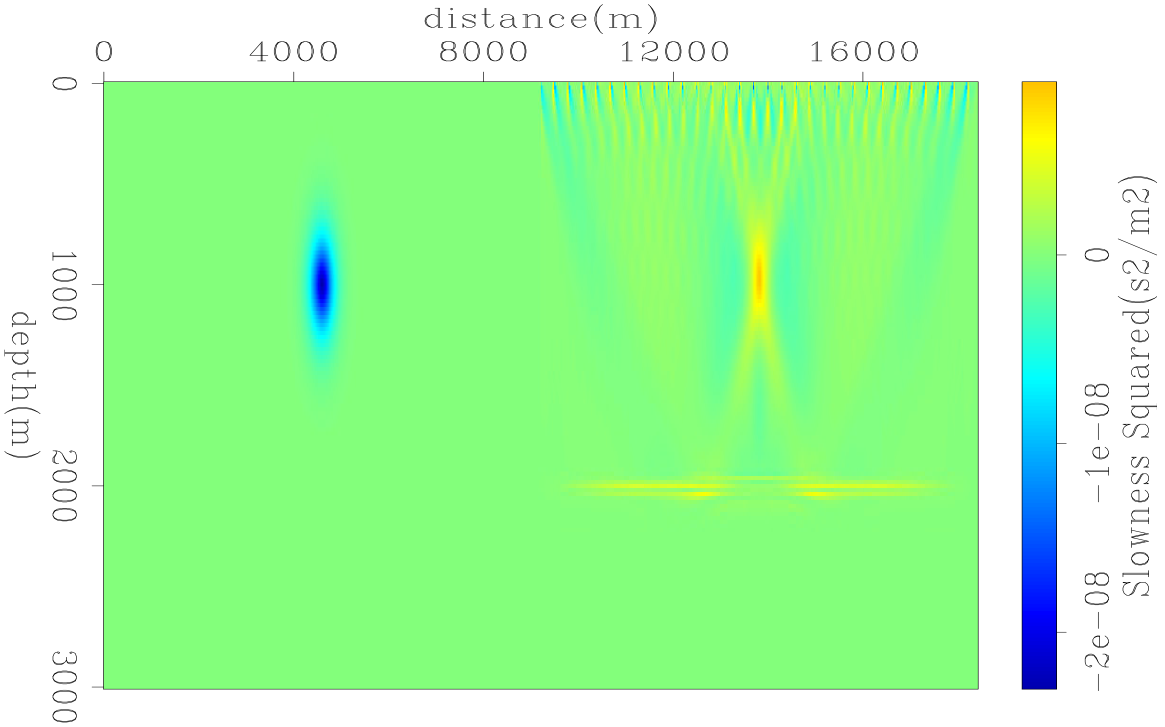

Figure 6: Perturbation in slowness squared, $\lambda = 0$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**].

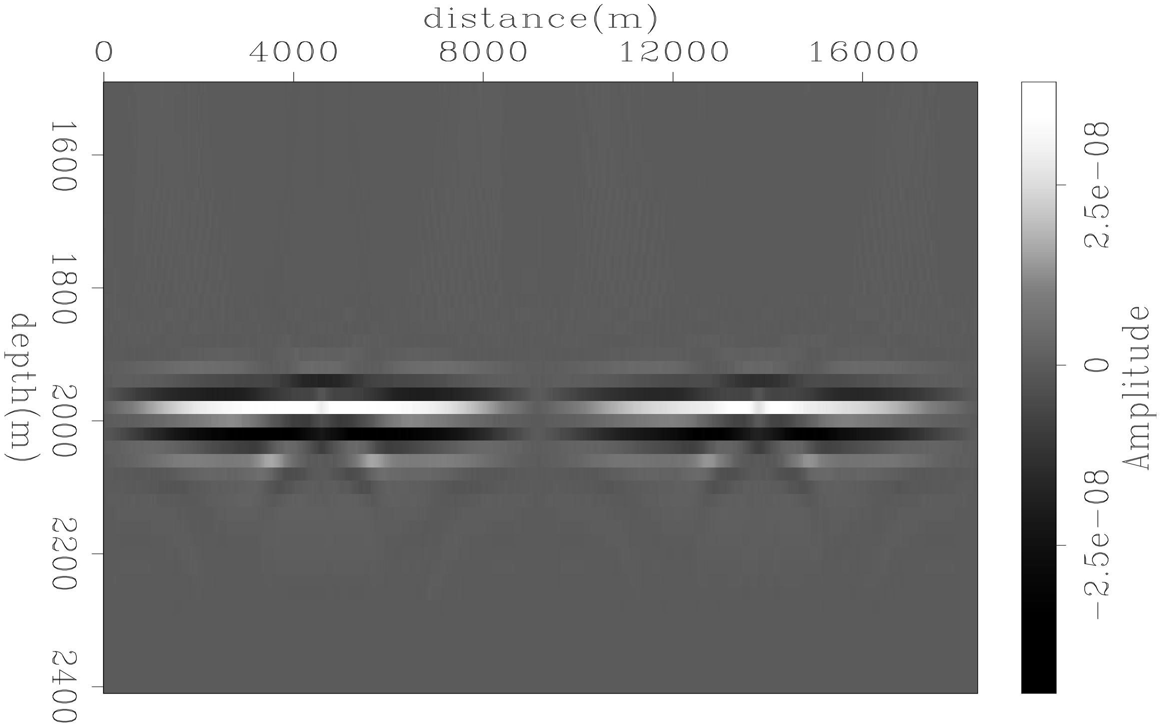

Figure 7: Reflectivity comparison, $\lambda = 0.5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**].

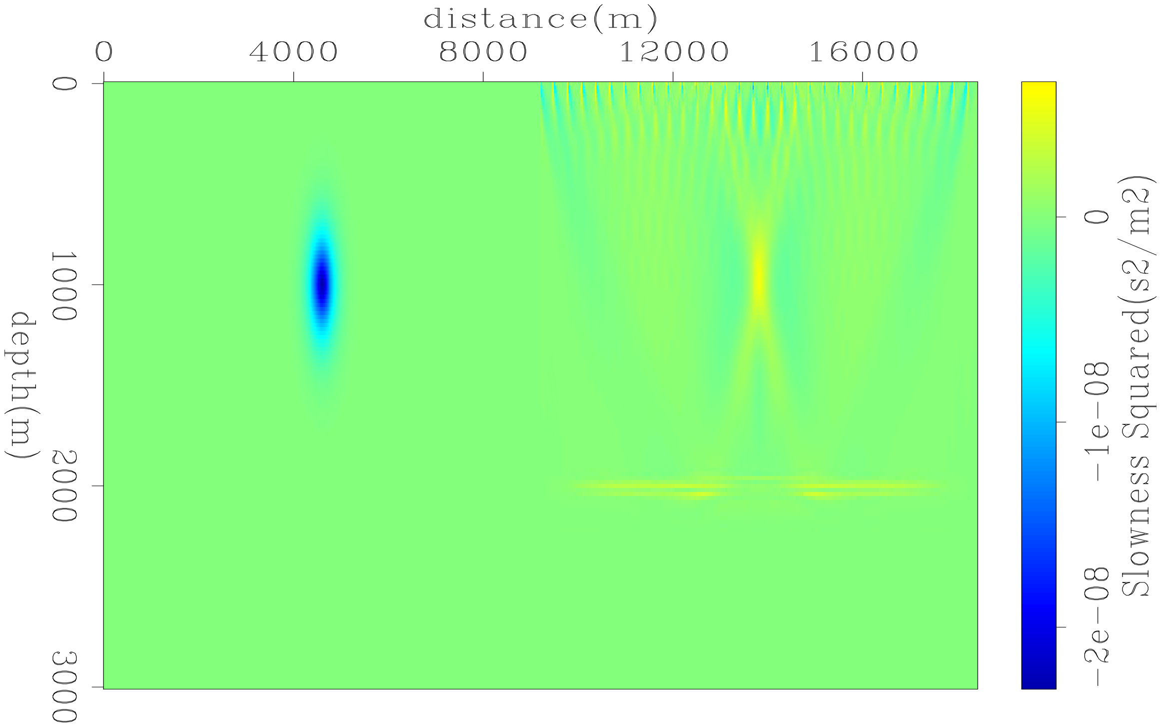

Figure 8: Perturbation in slowness squared, $\lambda = 0.5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**].

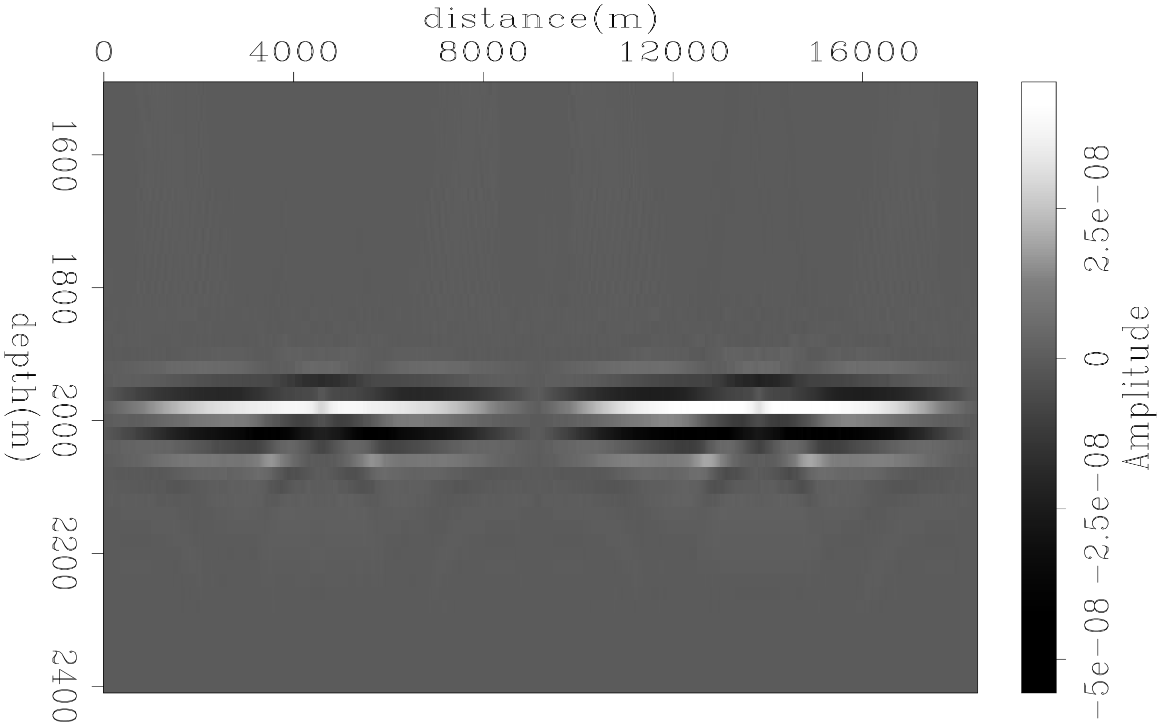

Figure 9: Reflectivity comparison, $\lambda = 1.0$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**].

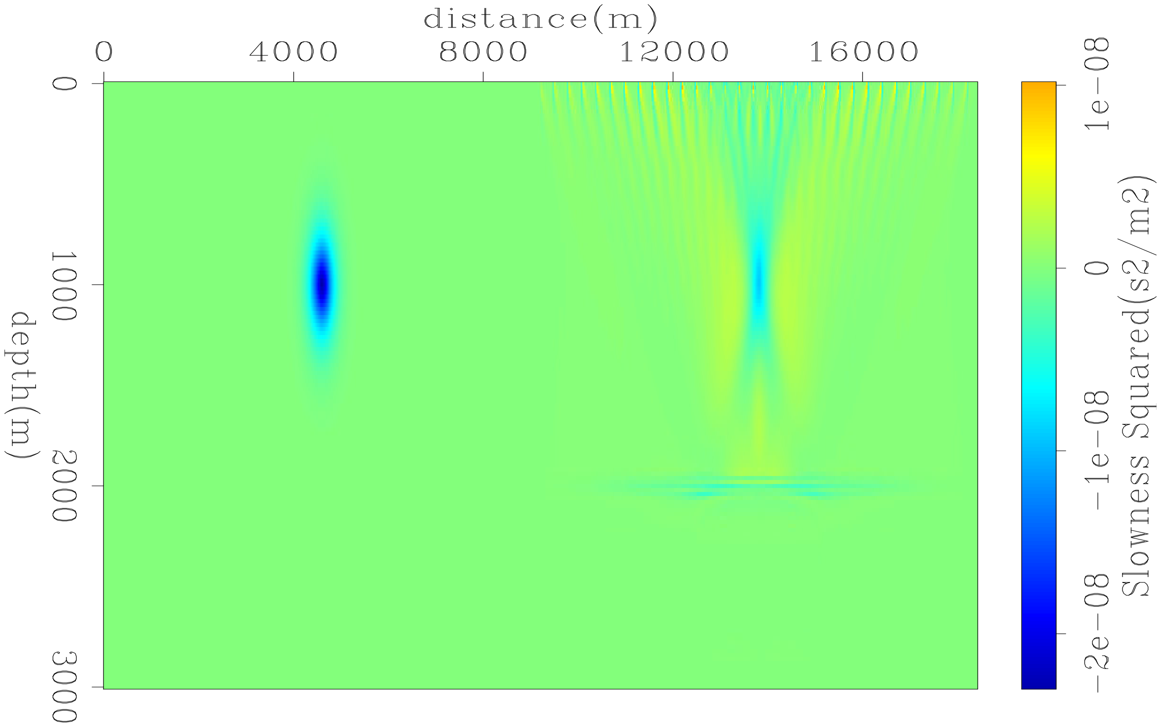

Figure 10: Perturbation in slowness squared, $\lambda = 1.0$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**].

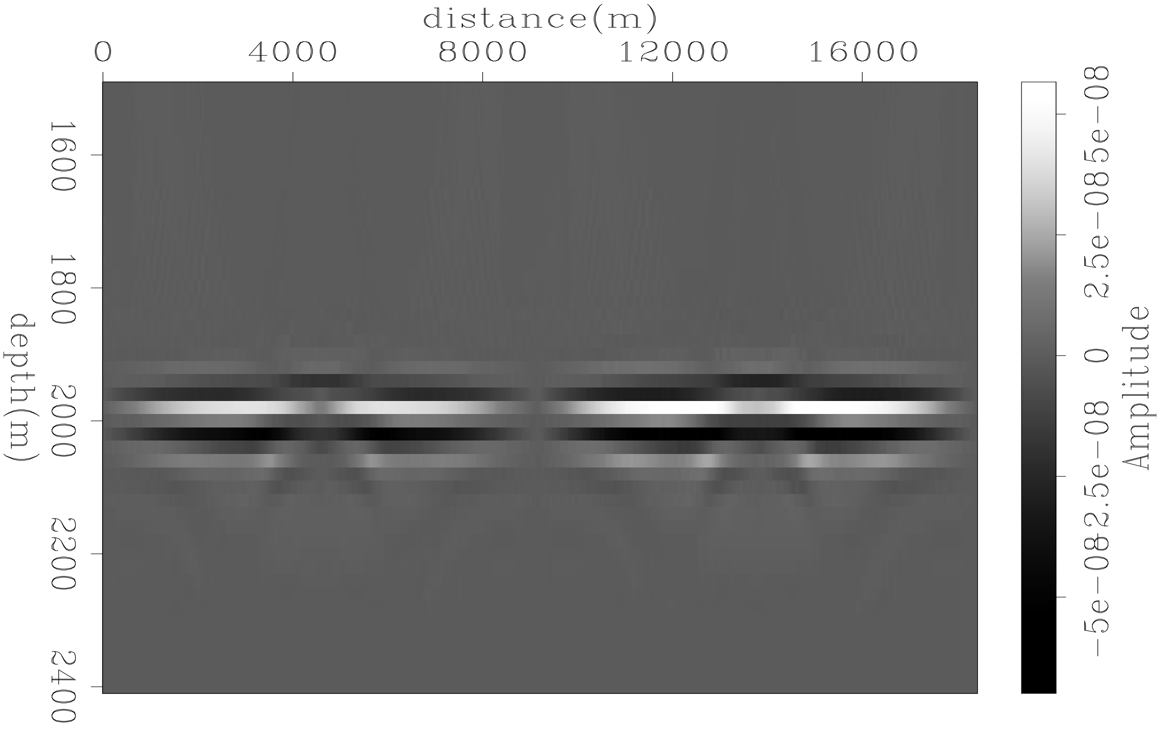

Figure 11: Reflectivity comparison, $\lambda = 1.0$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity. The number of iterations was increased to 16 [**CR**].

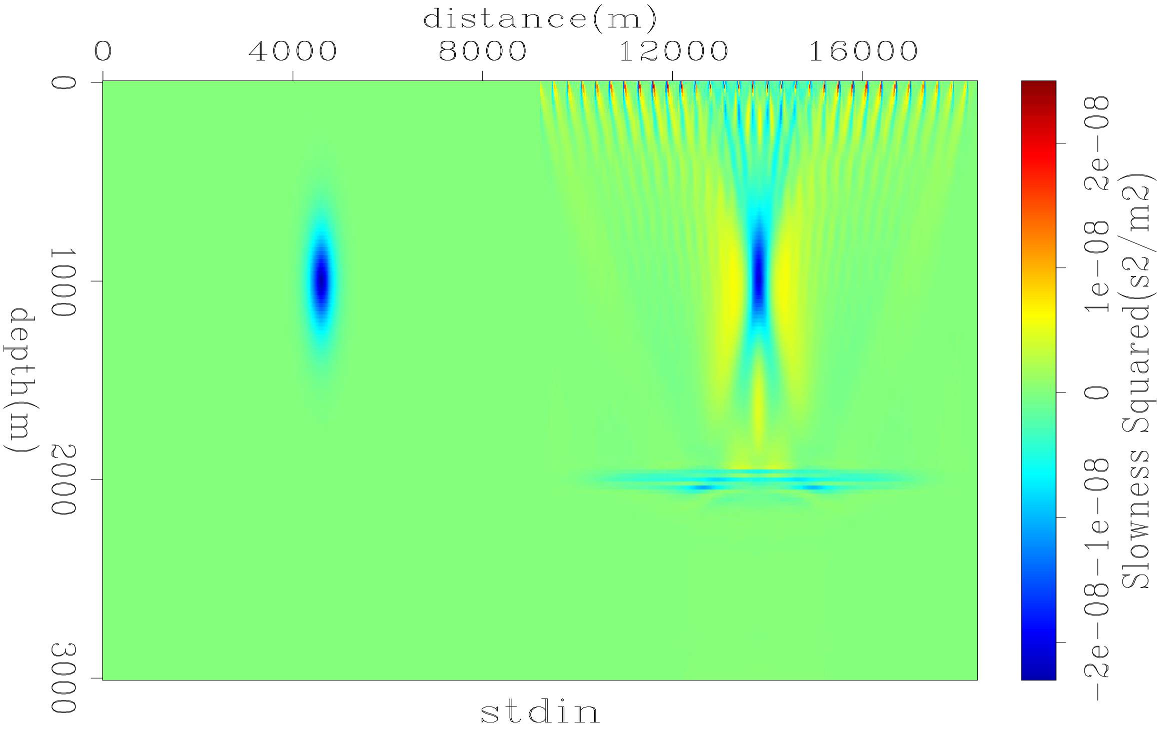

Figure 12: Perturbation in slowness squared, $\lambda = 1.0$. *Left*: True anomaly; *Right*: LWIVU perturbation. The number of iterations was increased to 16 [**CR**].

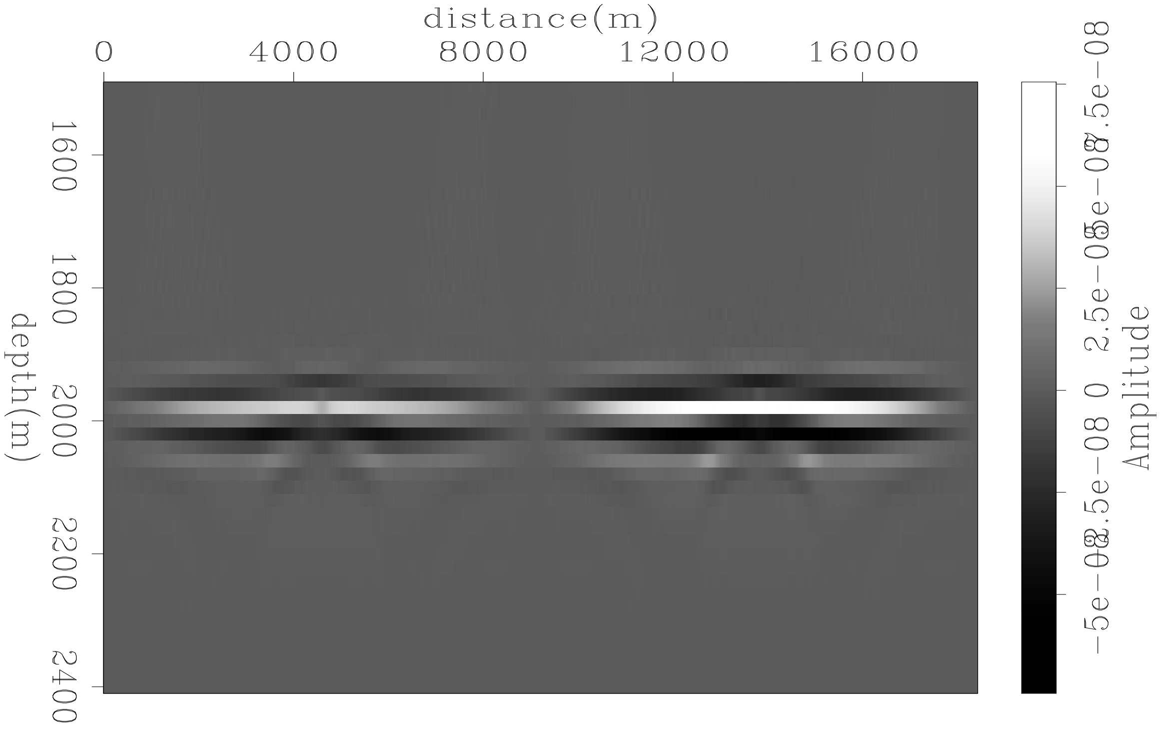

Figure 13: Reflectivity comparison, $\lambda = 1.5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**].

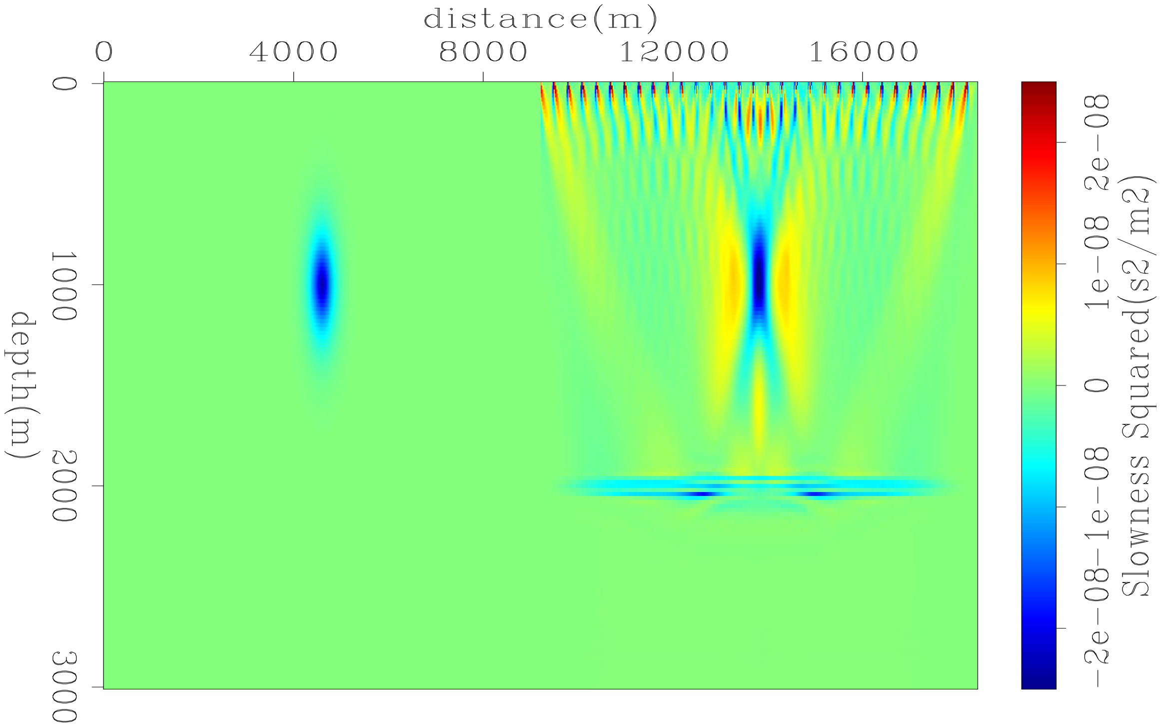

Figure 14: Perturbation in slowness squared, $\lambda = 1.5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**].

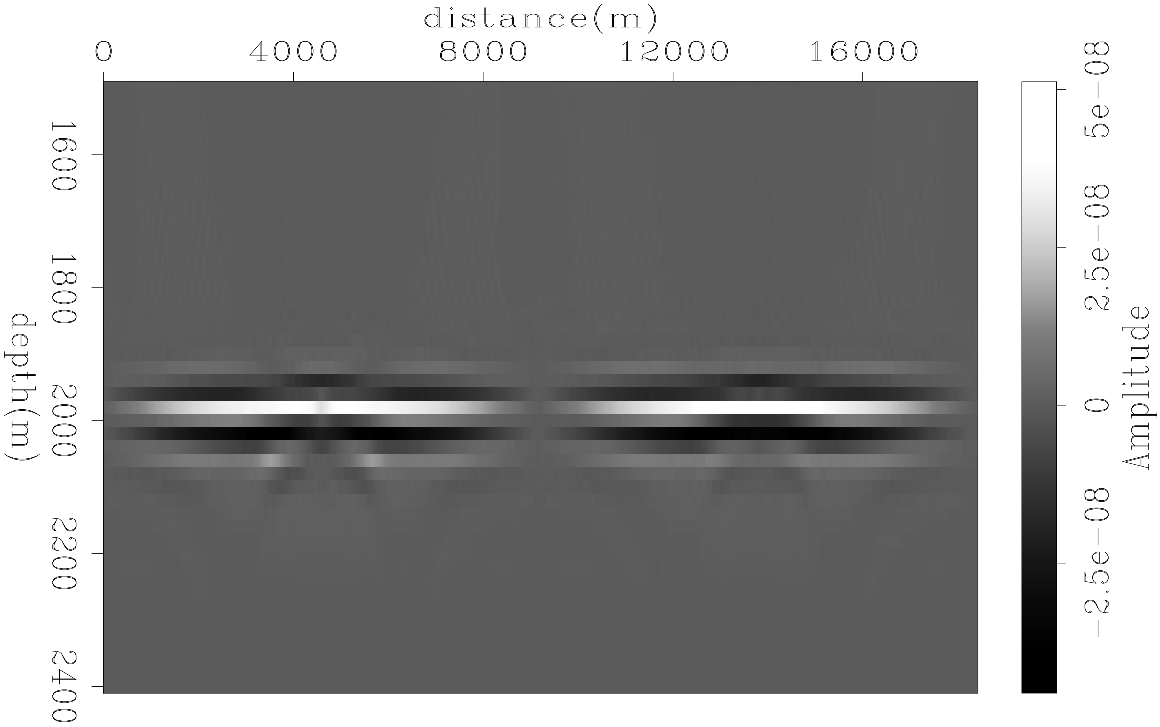

Figure 15: Reflectivity comparison, $\lambda = 2.5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**].

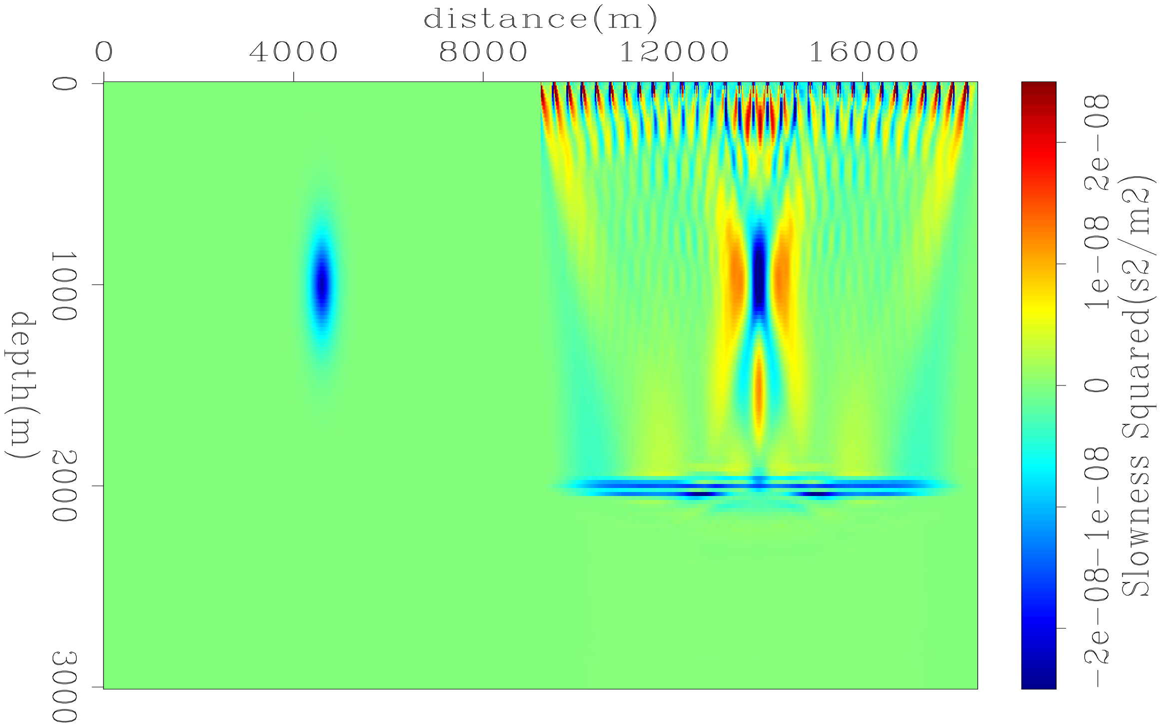

Figure 16: Perturbation in slowness squared, $\lambda = 2.5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**].

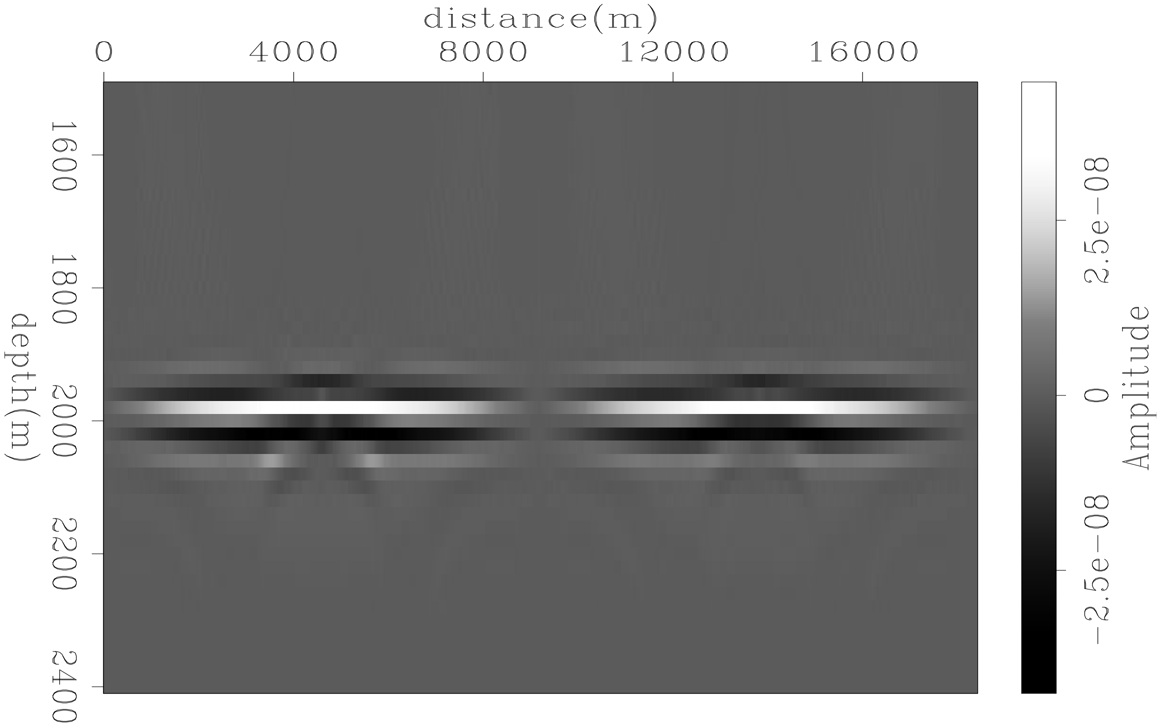

Figure 17: Reflectivity comparison, $\lambda = 2.5$, at 8 iterations. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**].

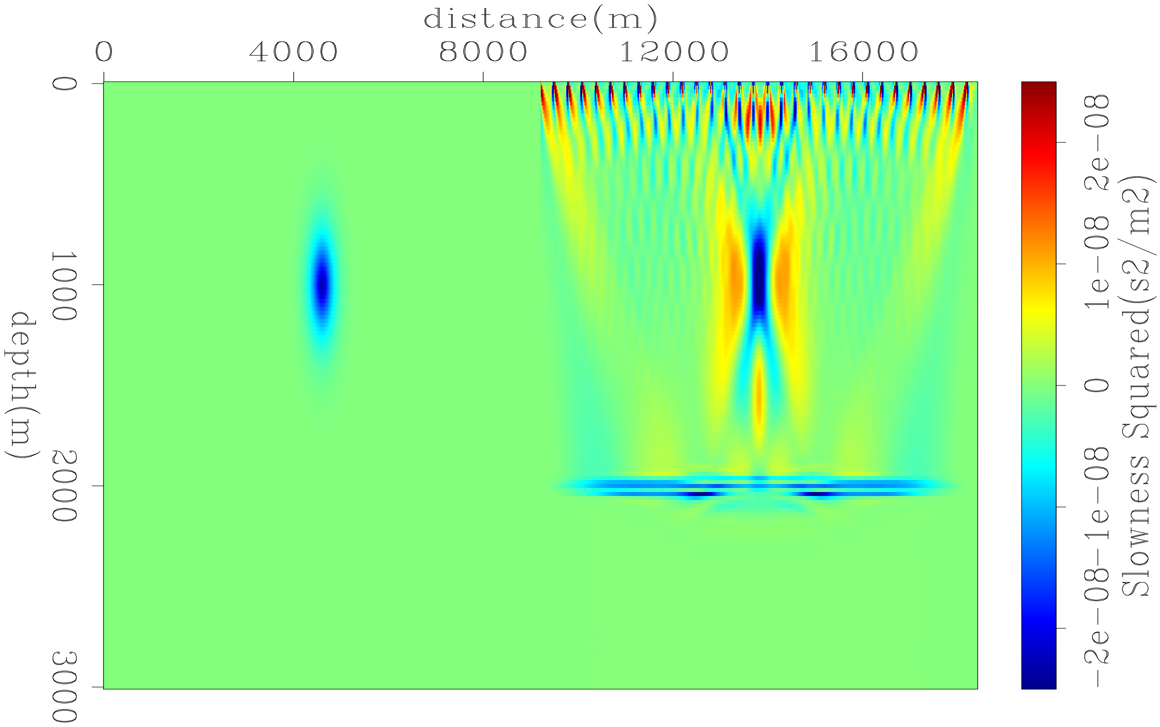

Figure 18: Perturbation in slowness squared, $\lambda = 2.5$, at 8 iterations. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**].

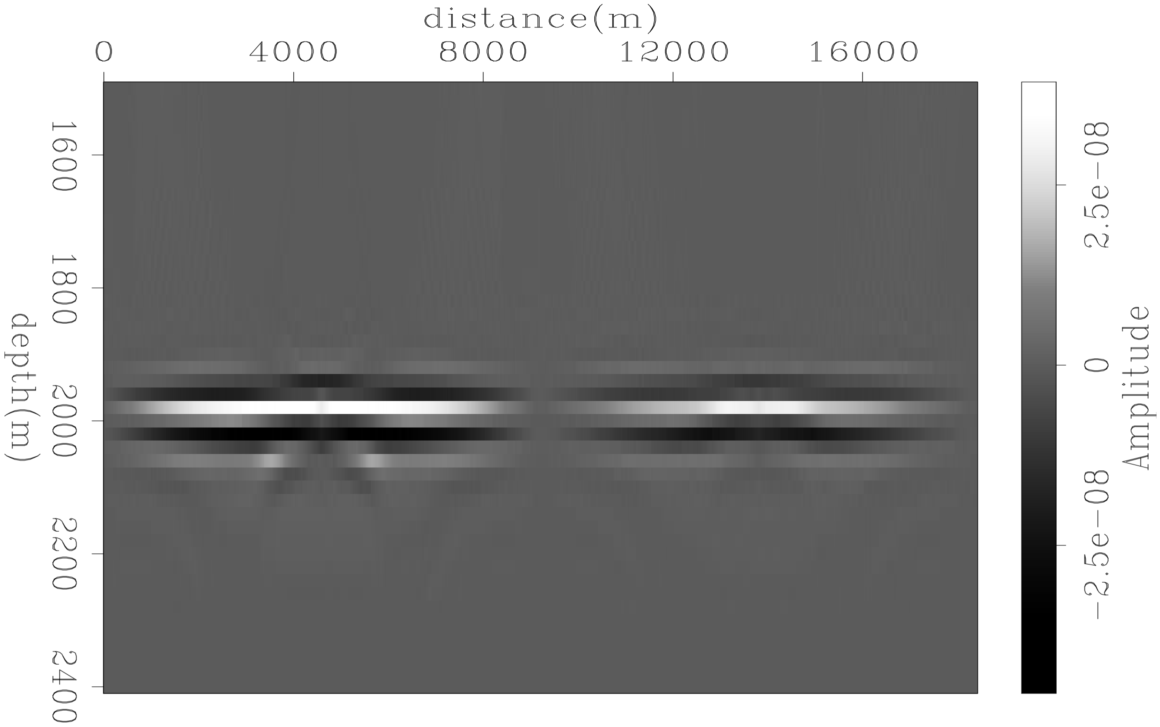

Figure 19: Reflectivity comparison, $\lambda = 5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**].

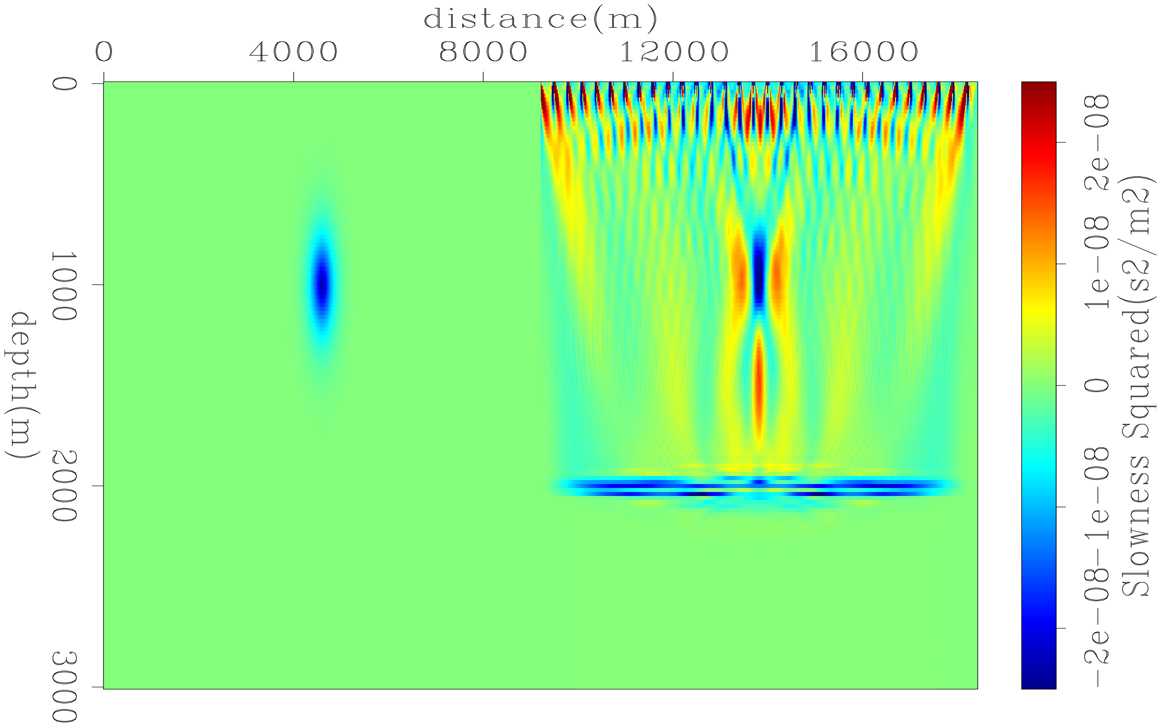

Figure 20: Perturbation in slowness squared, $\lambda = 5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**].

COMPUTATIONAL COST OF LWIVU

LWIVU demands more wavefield propagations than LWI because of the inclusion of WEMVA into the process. We compare LWI and LWIVU flowcharts to quantify them. The optimization scheme is the conjugate directions method (Appendix A). We assume reversible propagations by using random boundary conditions (Clapp, 2009, 2010), which massively reduce storage demand at the cost of extra-propagations. The arrows accompanying the source/receiver wavefield symbols signify forward propagation (right-pointing arrow) and backward propagation (left-pointing arrow) in time.

Figure 21 shows a flowchart of LWI performed in data space (in model space is trivial; LWI is virtually instantaneous once the Gauss-Newton Hessian has been pre-computed). We perform source wavefield and residual wavefield backward propagations in order to compute the gradient each iteration. In other words, the adjoint of Born modeling or RTM (a forward propagation of the source wavefield is required before the iterative scheme). Subsequent projection of the gradient into data space requires Born modeling. Thus, we perform forward propagation of the source wavefield and the source scattered wavefield, using the gradient as entry for reflectivity. The total number of propagations is $1+4n_{iter}$, where n_{iter} is the number of iterations.

Figure 22 shows a flowchart of LWIVU. We precompute the Gauss-Newton Hessian, e.g., using point-spread functions (Fletcher et al., 2016), and apply it on demand. Now the gradient consists of two members: one relates to the reflectivity and the other relates to the perturbation of the background model (see Appendix A for details). The former consists of applying the Gauss-Newton Hessian to the appropriate residual, not requiring wavefield propagations. The latter consists of applying the WEMVA Hessian to a linear combination of the residuals. Here WEMVA requires four backward wavefield propagations of the source wavefield and the receiver wavefield, among the souce scattered wavefield and the receiver scattered wavefield, using the residuals as entries for perturbations in the image (constituting the adjoing of WEMVA). For the projection of the gradient into data space we apply WEMVA again. We forward propagate the source wavefield and the receiver wavefield, among the source scattered wavefield and the receiver scattered wavefield using the appropriate member of the gradient as entry for perturbations of the background model (constituting forward WEMVA). Accounting for the initial forward propagation of the source wavefield, the total number of propagations is $1+8n_{iter}$, *i.e.* twice as much as in LWIVU, and excluding pre-computation of the Gauss-Newton Hessian.

From this analysis we recognize the computational demands as the main caveat of LWIVU. Nonetheless, we envision a situation where this method can be applicable. In reservoir characterization, it can be advantageous to perform LWI in model space in order to take advantage of amplitude variations with enhanced resolution, for better identification of oil-targets. We typically use small seismic volumes encompassing only the oil field of interest, so expensive iterative-based imaging methods become affordable. In case of small velocity errors impacting the amplitudes, we can implement LWIVU to rectify the reflectivity, without re-computing the Gauss-Newton Hessian.

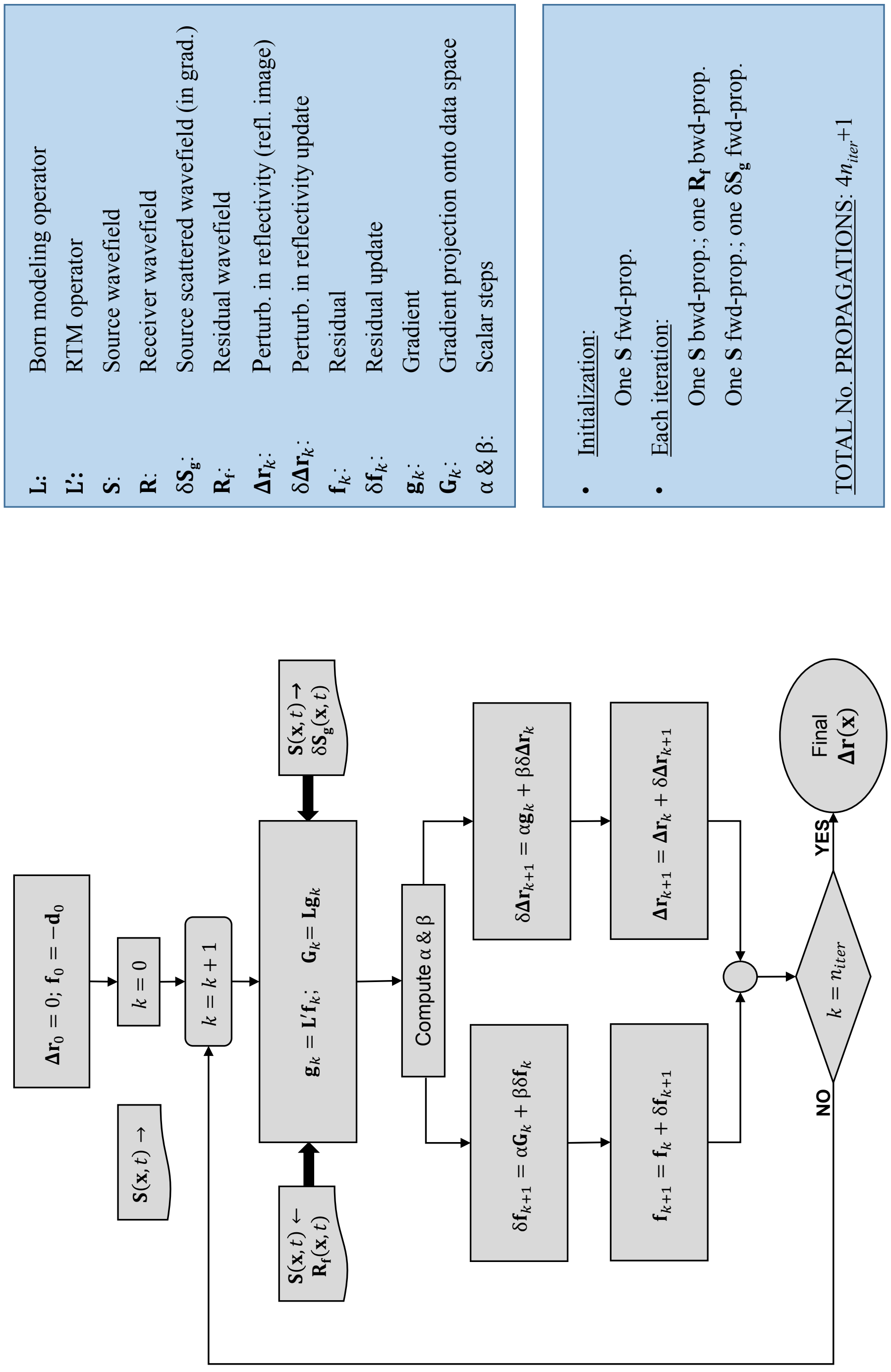

Figure 21: Flowchart of LWI in data space [**NR**].

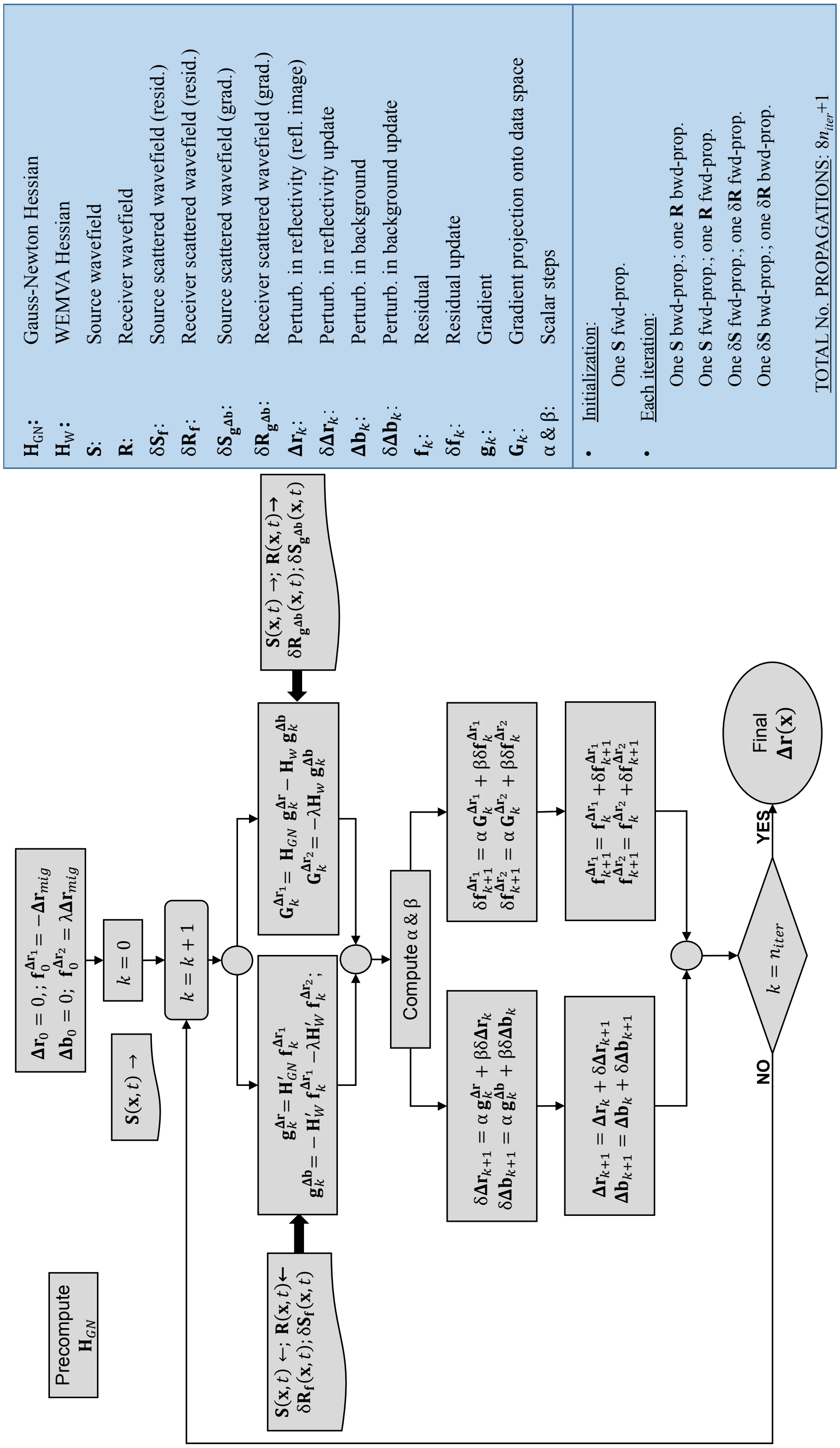

Figure 22: Flowchart of LWIVU [**NR**].

CONCLUSIONS

We revisited the LWIVU theory to make the improved image (after stacking power maximization) the target image for the reflectivity estimation. The synthetic examples demonstrate the potential of this method to correct amplitude inaccuracies derived from the cumulative effect of velocity (or slowness) errors. The main disadvantage is the computational demands, which are about twice compared to LWI, besides requiring the pre-computation of the Gauss-Newton Hessian. However, we foresee the applicability of LWIVU in detailed exploration works, such as reservoir characterization, where accurately addressing amplitude variations is vital for identification and delimitation of oil-bearing rock facies.

ACKNOWLEDGEMENT

We thank the SEP sponsors for their continuous support. Alejandro Cabrales thanks Petróleos Mexicanos for sponsoring his studies.

APPENDIX A: THE CONJUGATE DIRECTIONS METHOD AND LWIVU

The software that implements the gradient-based optimization is based on the conjugate directions method (Claerbout, 2014), a "cousin" of the more popular conjugate gradient method (Hestenes and Stiefel, 1952). In this Appendix I illustrate its implementation directly with LWIVU.

I initialize the method with an starting model. In linear problems it is often practical to set this model to zero. The model space consists of the perturbation of reflectivity and perturbation of the background, so the initial model is

$$\begin{bmatrix} \mathbf{\Delta r}_0 \\ \mathbf{\Delta b}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{A-1}$$

Then I compute the first residual, $\mathbf{f}_0$, by using the fitting goals in Equation 25 evaluated at the initial model:

$$\mathbf{f}_0 = \begin{bmatrix} \mathbf{f}_0^{\mathbf{\Delta r}_1} \\ \mathbf{f}_0^{\mathbf{\Delta r}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{GN}\mathbf{\Delta r}_0 - \mathbf{H}_W\mathbf{\Delta b}_0 - \mathbf{\Delta r}_{mig} \\ \lambda(\mathbf{H}_W\mathbf{\Delta b}_0 + \mathbf{\Delta r}_{mig}) \end{bmatrix} = \begin{bmatrix} -\mathbf{\Delta r}_{mig} \\ \lambda\mathbf{\Delta r}_{mig} \end{bmatrix} \tag{A-2}$$

The data space is the conventionally-migrated image space. Thus, the two components of the residual, $\mathbf{f}_0^{\mathbf{\Delta r}_1}$ and $\mathbf{f}_0^{\mathbf{\Delta r}_2}$, constitute migrated-like images.

Now I can set forth the iterative process, calculating the gradient at iteration k using the adjoind operator (Equation 28),

$$\mathbf{g}_k = \begin{bmatrix} \mathbf{g}_k^{\mathbf{\Delta r}} \\ \mathbf{g}_k^{\mathbf{\Delta b}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}'_{GN} & \mathbf{0} \\ -\mathbf{H}'_W & -\lambda\mathbf{H}'_W \end{bmatrix} \begin{bmatrix} \mathbf{f}_k^{\mathbf{\Delta r}_1} \\ \mathbf{f}_k^{\mathbf{\Delta r}_2} \end{bmatrix} \tag{A-3}$$

and then projecting such gradient onto the data space using the forward operator (Equation 30),

$$\mathbf{G}_k = \begin{bmatrix} \mathbf{G}_k^{\mathbf{\Delta r}_1} \\ \mathbf{G}_k^{\mathbf{\Delta r}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{GN} & -\mathbf{H}_W \\ \mathbf{0} & -\lambda \mathbf{H}_W \end{bmatrix} \begin{bmatrix} \mathbf{g}_k^{\mathbf{\Delta r}} \\ \mathbf{g}_k^{\mathbf{\Delta b}} \end{bmatrix}. \tag{A-4}$$

What follows next is determining the model update and the residual update. They are given respectively by

$$\begin{bmatrix} \delta \mathbf{\Delta r}_{k+1} \\ \delta \mathbf{\Delta b}_{k+1} \end{bmatrix} = \alpha \begin{bmatrix} \mathbf{g}_k^{\mathbf{\Delta r}} \\ \mathbf{g}_k^{\mathbf{\Delta b}} \end{bmatrix} + \beta \begin{bmatrix} \delta \mathbf{\Delta r}_k \\ \delta \mathbf{\Delta b}_k \end{bmatrix} \tag{A-5}$$

and

$$\begin{bmatrix} \delta \mathbf{f}_{k+1}^{\mathbf{\Delta r}_1} \\ \delta \mathbf{f}_{k+1}^{\mathbf{\Delta r}_2} \end{bmatrix} = \alpha \begin{bmatrix} \mathbf{G}_k^{\mathbf{\Delta r}_1} \\ \mathbf{G}_k^{\mathbf{\Delta r}_2} \end{bmatrix} + \beta \begin{bmatrix} \delta \mathbf{f}_k^{\mathbf{\Delta r}_1} \\ \delta \mathbf{f}_k^{\mathbf{\Delta r}_2} \end{bmatrix}, \tag{A-6}$$

where α and β are parameters that define the search plane that minimizes the new residual (Claerbout, 2014). Finally, I update the model and the residual:

$$\begin{bmatrix} \mathbf{\Delta r}_{k+1} \\ \mathbf{\Delta b}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Delta r}_k \\ \mathbf{\Delta b}_k \end{bmatrix} + \begin{bmatrix} \delta \mathbf{\Delta r}_{k+1} \\ \delta \mathbf{\Delta b}_{k+1} \end{bmatrix}, \tag{A-7}$$

$$\begin{bmatrix} \mathbf{f}_{k+1}^{\mathbf{\Delta r}_1} \\ \mathbf{f}_{k+1}^{\mathbf{\Delta r}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_k^{\mathbf{\Delta r}_1} \\ \mathbf{f}_k^{\mathbf{\Delta r}_2} \end{bmatrix} + \begin{bmatrix} \delta \mathbf{f}_{k+1}^{\mathbf{\Delta r}_1} \\ \delta \mathbf{f}_{k+1}^{\mathbf{\Delta r}_2} \end{bmatrix}, \tag{A-8}$$

and proceed with the next iteration.

APPENDIX B: WEMVA INVERSION USING STACKING POWER MAXIMIZATION

In this Appendix I derive the WEMVA inversion process using the maximization of the stacking power as a focusing operator. This derivation is important to have a better understanding of how LWIVU was derived, for maximizing a function via minimizing its negation is tricky. Thus, I explain with special thoroughness through this section.

The general expression for the WEMVA misfit function is (Biondi, 2006)

$$\Phi(\mathbf{s}^2) = \frac{1}{2} \| \mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \mathbf{F}\left[\mathbf{\Delta r}_{mig}(\mathbf{s}^2)\right] \|_2^2, \tag{B-1}$$

where $\mathbf{s}^2$ is the slowness squared field or background model, and $\mathbf{F}$ is a focusing operator applied to the migrated image. For the sake of consistency I keep the notation for the migrated image, $\mathbf{\Delta r}_{mig}$, the same as in the theoretical section.

The focusing operator enforces the correction of the migrated image *without changing the background model.* Such corrected image will become the *target image* that the inversion process fits, now updating the background model. There are several focusing operators. The most popular are perhaps the *differential semblance optimization* (DSO) (Symes and Carazzone, 1991) and the maximization of the stacking power. The former enforces focusing by flattening the angle-domain common-image gathers (ADCIG), whereas the latter seeks for maximum power of the stack section, presumably occurring when offset-domain common-image gathers (ODCIG) focus, an it is equivalent to flattening ADCIG. Maximazing the stacking power is known to suffer from cycle skipping when velocity errors are big, but in LWIVU we assume that velocity inaccuracies are rather small, therefore it becomes applicable.

Maximization of the stacking power is achieved using $\mathbf{F} = \mathbf{I} + \mathbf{S}$ as a focusing operator, where $\mathbf{I}$ is the identity operator and $\mathbf{S}$ is the stacking operator. Substituting into Equation B-1, we obtain

$$\Phi(\mathbf{s}^2) = \frac{1}{2}\|\mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \left[\mathbf{\Delta r}_{mig}(\mathbf{s}^2) + \mathbf{S}\mathbf{\Delta r}_{mig}(\mathbf{s}^2)\right]\|_2^2 = \frac{1}{2}\| - \mathbf{S}\mathbf{\Delta r}_{mig}(\mathbf{s}^2)\big]\|_2^2 \quad \text{(B-2)}$$

The minus sign is lost if we simply perform the algebra, yielding an ordinary minimization. In order to achieve the maximization we move the negative sign out of the absolute value (e.g. Tang, 2011a),

$$\Phi(\mathbf{s}^2) = \frac{1}{2}\|\mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \left[\mathbf{\Delta r}_{mig}(\mathbf{s}^2) + \mathbf{S}\mathbf{\Delta r}_{mig}(\mathbf{s}^2)\right]\|_2^2 = -\frac{1}{2}\|\mathbf{S}\mathbf{\Delta r}_{mig}(\mathbf{s}^2)\big]\|_2^2 \quad \text{(B-3)}$$

In the case of zero subsurface offset, the stacking operator becomes the identity operator, thus

$$\Phi(\mathbf{s}^2) = -\frac{1}{2}\|\mathbf{\Delta r}_{mig}(\mathbf{s}^2)\big]\|_2^2 \quad \text{(B-4)}$$

We can linearize this problem by assuming that $\mathbf{s}^2 = \mathbf{s}_0^2 + \mathbf{\Delta s}^2$, where $\mathbf{s}_0^2$ is the initial slowness squared model (analogous to the most background model), and $\mathbf{\Delta s}^2$ is a perturbation in such initial model. Thus, I can expand the migrated image around $\mathbf{s}_0^2$, with $\mathbf{s}^2 = \mathbf{s}_0^2 + \mathbf{\Delta s}^2$ for a small perturbation $\mathbf{\Delta s}^2$,

$$\mathbf{\Delta r}_{mig}(\mathbf{s}^2) \approx \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \left[\frac{\partial \mathbf{\Delta r}_{mig}(\mathbf{s}^2)}{\partial \mathbf{s}^2}|_{\mathbf{s}^2=\mathbf{s}_0^2}\right]\mathbf{\Delta s}^2 = \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W}\mathbf{\Delta s}^2, \quad \text{(B-5)}$$

where $\mathbf{\Delta r}_{mig}(\mathbf{s}_0^2)$ is the image migrated with $\mathbf{s}_0^2$, and The derivative in Equation B-7 constitutes the WEMVA operator, $\mathbf{W}$ (equivalent to the WEMVA Hessian, $\mathbf{H}_W$). Note from the expansion in Equation B-5 that the forward WEMVA operator constitutes a linear operator that relates perturbations in the background slowness squared to perturbations in the image:

$$\mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) = \mathbf{\Delta}(\mathbf{\Delta r}_{mig}) \approx \mathbf{W}\mathbf{\Delta s}^2. \quad \text{(B-6)}$$

Now I obtain the gradient of the misfit function (Equation B-4),

$$\begin{aligned}\nabla\Phi(\mathbf{s}^2) &= -\left[\frac{\partial \mathbf{\Delta r}_{mig}(\mathbf{s}^2)}{\partial \mathbf{s}^2}\bigg|_{\mathbf{s}^2=\mathbf{s}_0^2}\right]'[\mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W}\mathbf{\Delta s}^2] \\ &= -\mathbf{W}'[\mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W}\mathbf{\Delta s}^2], \end{aligned} \quad \text{(B-7)}$$

Finally, I can initialize the conjugate directions method by setting the initial model and the initial residual, the latter evaluating the fitting goal on the initial model (see Appendix A),

$$\mathbf{\Delta s}_0^2 = \mathbf{0}; \quad \mathbf{f}_0 = \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W}\mathbf{\Delta s}_0^2 = \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2). \tag{B-8}$$

The gradient is defined with the *negative* of adjoint operator (Equation B-7) applied to the residuals,

$$\mathbf{g}_k = -\mathbf{W}'\mathbf{f}_k. \tag{B-9}$$

and it is projected onto the data space with the *negative* of the forward operator

$$\mathbf{G}_k = -\mathbf{W}\mathbf{g}_k. \tag{B-10}$$

Note that the negative signs accompanying the forward and the adjoint WEMVA operators in Equations B-9 and B-10 are also utilized in the WEMVA Hessian for the LWIVU operator (see Appendix A).

REFERENCES

Almomin, A., 2013, Accurate implementation of two-way wave-equation operators: SEP-Report, **149**, 281–288.

Barnier, G. and A. Almomin, 2014, Tutorial on two-way wave equation operators for acoustic, isotropic, constant-density media: SEP-Report, **155**, 23–56.

Biondi, B., 2006, 3-D seismic imaging: Society of Exploration Geophysicists.

Biondi, B. and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, no. 3, WA129–WA140.

Biondi, B., E. Biondi, M. Maharramov, and Y. Ma, 2015, Dissection of the full-waveform inversion hessian: SEP-Report, **160**, 19–38.

Cabrales-Vargas, A., B. Biondi, and R. Clapp, 2016, Linearized waveform inversion with (small) velocity updating : SEP-Report, **163**, 189–196.

Claerbout, J., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications, Inc.

———, 2014, Geophysical image estimation by example: Jon Claerbout.

Clapp, M., 2005, Imaging under salt: Illumination compensation by regularized inversion: PhD thesis, Stanford University.

Clapp, R., 2009, Reverse time migration with random boundaries: SEG Technical Program Expanded Abstracts, 2809–2813.

Clapp, R. G., 2010, More fun with random boundaries: SEP-Report, **140**, 95–102.

Dai, W., C. Boonyasiriwat, and G. Schuster, 2010, 3D multi-source least-squares reverse time migration: SEG Technical Program Expanded Abstracts, 3120–3124.

Fichtner, A., 2011, Full seismic waveform modelling and inversion: Springer Science & Business Media.

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.

Guitton, A., A. Valenciano, D. Bevc, and J. Claerbout, 2007, Smoothing imaging conditions for shot-profile migration: Geophysics, **72**, no. 3, S149–S154.

Hestenes, M. and E. Stiefel, 1952, Methods of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**, no. 6, 409–436.

Luo, S. and D. Hale, 2014, Reverse time migration in the presence of velocity errors: Geophysics, **79**, no. 4, S153–S161.

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

Rocha, D., N. Tanushev, and P. Sava, 2016, Acoustic wavefield imaging using the energy norm: Geophysics, **81**, no. 4, S151–S163.

Ronen, S. and C. Liner, 2000, Least-squares DMO and migration: Geophysics, **65**, 1364–1371.

Symes, W. and J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Tang, Y., 2011a, Fast automatic wave-equation migration velocity analysis using encoded simultaneous sources: SEP-Report, **145**, 27–46.

———, 2011b, Imaging and velocity analysis by target-oriented wavefield inversion: PhD thesis, Stanford University.

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

Valenciano, A., B. Biondi, and R. Clapp, 2009, Imaging by target-oriented wave-equation inversion: Geophysics, **74**, no. 6, WCA109–WCA120.

Valenciano, A. A., 2008, Imaging by wave-equation inversion: PhD thesis, Stanford University.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26.