Extended linearized waveform inversion with velocity updating

Alejandro Cabrales-Vargas

ABSTRACT

In this report I recast the objective function of linearized waveform inversion with velocity updating into a subsurface offset extended version. This approach prevents the corresponding Gauss-Newton Hessian from becoming non-positive definite. Both theory and numerical examples appear to support this asseveration, although some work is still required for proper precondition of the extended wave-equation migration velocity analysis operator. Additionally, the new implementation demands more computational resources as a consequence of the extension.

INTRODUCTION

Linearized waveform inversion with velocity updated (LWIVU) consists of coupling the full waveform inversion (FWI) Gauss-Newton Hessian with the wave-equation migration velocity analysis (WEMVA) operator, in an effort to obtain a better estimation of the subsurface reflectivity. The original method employs maximization of the WEMVA-updated migration image power as a constraint:

$$\Phi(\mathbf{r},\mathbf{\Delta b}) = \frac{1}{2}\|\mathbf{Hr}-\mathbf{W\Delta b}-\mathbf{r}_m\|_2^2 - \frac{\lambda^2}{2}\|\mathbf{W\Delta b}+\mathbf{r}_m^0\|_2^2. \tag{1}$$

Here $\mathbf{H}$ is the FWI Gauss-Newton Hessian, $\mathbf{W}$ is the WEMVA operator, $\mathbf{r}$ is the reflectivity, $\mathbf{\Delta b}$ is the perturbation in the background slowness squared, $\mathbf{r}_m^0 = \mathbf{r}_m(\mathbf{b}_0)$ is the migration image using the most background model, $\mathbf{b}_0$, and λ is a trade-off factor that controls the influence of the second functional in Equation 1. The minus sign before $\frac{\lambda^2}{2}$ allows maximization of the image power by minimizing its negative value.

In the preparation of the numerical examples of the previous report (Cabrales-Vargas, 2018) I found that the solution of Equation 1 was not unique. In fact, the result depended on the solving method (*e.g.*, conjugate directions, steepest descent, or a combination of both). Therefore, in this report I start with the analysis of the current objective function (Equation 1). Next, I recast LWIVU to an extended version, using the differential semblance optimization (DSO) operator as part of the constraint function, discussing the new challenges posed and the proposed solutions. Finally, I comment on the future steps to be followed.

THEORY

Original approach: Regularize maximizing the power of the migration image

The gradient of the objective function in Equation 1 is obtained by deriving it with respect to the parameters $\mathbf{r}$ and $\mathbf{\Delta b}$:

$$\nabla\Phi = \begin{bmatrix} \nabla_{\mathbf{r}}\Phi \\ \nabla_{\mathbf{\Delta b}}\Phi \end{bmatrix} = \begin{bmatrix} \mathbf{H}'\big(\mathbf{Hr} - \mathbf{W\Delta b} - \mathbf{r}_m^0\big) \\ -\mathbf{W}'\big(\mathbf{Hr} - \mathbf{W\Delta b} - \mathbf{r}_m^0\big) - \lambda^2\mathbf{W}'\big(\mathbf{W\Delta b} + \mathbf{r}_m^0\big) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{H}' & \mathbf{0} \\ -\mathbf{W}' & -\lambda\mathbf{W}' \end{bmatrix} \begin{bmatrix} \mathbf{Hr} - \mathbf{W\Delta b} - \mathbf{r}_m^0 \\ \lambda\big(\mathbf{W\Delta b} + \mathbf{r}_m^0\big) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{H}' & \mathbf{0} \\ -\mathbf{W}' & -\lambda\mathbf{W}' \end{bmatrix} \left\{ \begin{bmatrix} \mathbf{H} & -\mathbf{W} \\ \mathbf{0} & \lambda\mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{\Delta b} \end{bmatrix} - \begin{bmatrix} \mathbf{r}_m^0 \\ -\lambda\mathbf{r}_m^0 \end{bmatrix} \right\} \tag{2}$$

The last row in Equation 2 can be compared with the gradient of a generic misfit function $\Phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{Lm} - \mathbf{d}\|_2^2$, given as

$$\nabla\Phi = \mathbf{L}'\big(\mathbf{Lm} - \mathbf{d}\big). \tag{3}$$

From Equation 2 we identify

$$\mathbf{L} = \begin{bmatrix} \mathbf{H} & -\mathbf{W} \\ \mathbf{0} & \lambda\mathbf{W} \end{bmatrix}, \tag{4}$$

however, we also notice that

$$\mathbf{L}' \neq \begin{bmatrix} \mathbf{H}' & \mathbf{0} \\ -\mathbf{W}' & -\lambda\mathbf{W}' \end{bmatrix}, \tag{5}$$

The discrepancy between the signs of the terms $\lambda\mathbf{W}$ and $\lambda\mathbf{W}'$ in Equations 4 and 5 prevents the LWIVU operators from becoming adjoint of each other, hence Equation 2 cannot be expressed as the generic form (Equation 3). In addition, if we cascade such operators to build the LWIVU Gauss-Newton Hessian, $\mathcal{H}$, we obtain

$$\mathcal{H} = \begin{bmatrix} \mathbf{H}' & \mathbf{0} \\ -\mathbf{W}' & -\lambda\mathbf{W}' \end{bmatrix} \begin{bmatrix} \mathbf{H} & -\mathbf{W} \\ \mathbf{0} & \lambda\mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{H}'\mathbf{H} & -\mathbf{H}'\mathbf{W} \\ -\mathbf{W}'\mathbf{H} & (1-\lambda^2)\mathbf{W}'\mathbf{W} \end{bmatrix}. \tag{6}$$

The term $1 - \lambda^2$ plays a crucial role in the potential ill-posedness of the method. If $\lambda > 1$ the problem is no longer positive definite. As a consequence, gradient-based methods will not converge to a minimum. In particular, if $\lambda \approx 1$, the model null space becomes prominent, thus jeopardizing the reliability of the solution. Therefore, albeit not necessarily invalidating the method, this limitation seriously restricts our freedom of choice for the values of λ. This fact explains why in early examples the method appeared to be promising, but failed as bigger values of λ were required.

Alternative approach: Regularizing with differential semblance optimization

As aforementioned, the main problem of non-extended LWIVU is the risk of ill-posedness due to coupling two functionals where one is minimized whilst the other maximized. The obvious solution is replacing the latter with a functional that requires minimization. One way of accomplishing this task is employing the DSO operator (Symes and Carazzone, 1991) in the subsurface offset domain. We recast LWIVU as

$$\Phi(\hat{\mathbf{r}}, \mathbf{\Delta b}) = \frac{1}{2}\|\hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{\Delta b} - \hat{\mathbf{r}}_m^0\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{D}\big(\hat{\mathbf{W}}\mathbf{\Delta b} + \hat{\mathbf{r}}_m^0\big)\|_2^2, \tag{7}$$

where $\mathbf{D}$ is the DSO operator. The hat "$\wedge$" is used to represent both operators and parameters that are extended in subsurface offset. Note that the perturbation in the background, $\mathbf{\Delta b}$, remains non-extended. On the other hand, the DSO operator is extended by definition.

One important change introduced with Equation 7 is the plus sign in the second functional. It translates into the minimization of both members. As a consequence, when we obtain the gradient,

$$\begin{aligned}
\nabla\Phi = \begin{bmatrix} \nabla_{\hat{\mathbf{r}}}\Phi \\ \nabla_{\mathbf{\Delta b}}\Phi \end{bmatrix} &= \begin{bmatrix} \hat{\mathbf{H}}'\big(\hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{\Delta b} - \hat{\mathbf{r}}_m^0\big) \\ -\hat{\mathbf{W}}'\big(\hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{\Delta b} - \hat{\mathbf{r}}_m^0\big) + \lambda^2\hat{\mathbf{W}}'\mathbf{D}'\mathbf{D}\big(\hat{\mathbf{W}}\mathbf{\Delta b} + \hat{\mathbf{r}}_m^0\big) \end{bmatrix} \\
&= \begin{bmatrix} \hat{\mathbf{H}}' & \mathbf{0} \\ -\hat{\mathbf{W}}' & \lambda\hat{\mathbf{W}}'\mathbf{D}' \end{bmatrix} \begin{bmatrix} \hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{\Delta b} - \hat{\mathbf{r}}_m^0 \\ \lambda\mathbf{D}\big(\hat{\mathbf{W}}\mathbf{\Delta b} + \hat{\mathbf{r}}_m^0\big) \end{bmatrix} \\
&= \begin{bmatrix} \hat{\mathbf{H}}' & \mathbf{0} \\ -\hat{\mathbf{W}}' & \lambda\hat{\mathbf{W}}'\mathbf{D}' \end{bmatrix} \left\{ \begin{bmatrix} \hat{\mathbf{H}} & -\hat{\mathbf{W}} \\ \mathbf{0} & \lambda\mathbf{D}\hat{\mathbf{W}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{r}} \\ \mathbf{\Delta b} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{r}}_m^0 \\ -\lambda\mathbf{D}\hat{\mathbf{r}}_m^0 \end{bmatrix} \right\},
\end{aligned} \tag{8}$$

we can identify the forward and adjoint operators strightforwardly as

$$\mathbf{L} = \begin{bmatrix} \hat{\mathbf{H}} & -\hat{\mathbf{W}} \\ \mathbf{0} & \lambda\mathbf{D}\hat{\mathbf{W}} \end{bmatrix}, \quad \mathbf{L}' = \begin{bmatrix} \hat{\mathbf{H}}' & \mathbf{0} \\ -\hat{\mathbf{W}}' & \lambda\hat{\mathbf{W}}'\mathbf{D}' \end{bmatrix}, \tag{9}$$

and therefore, the extended LWIVU Gauss-Newton Hessian becomes

$$\hat{\mathcal{H}} = \begin{bmatrix} \hat{\mathbf{H}}' & \mathbf{0} \\ -\hat{\mathbf{W}}' & \lambda\hat{\mathbf{W}}'\mathbf{D}' \end{bmatrix} \begin{bmatrix} \hat{\mathbf{H}} & -\hat{\mathbf{W}} \\ \mathbf{0} & \lambda\mathbf{D}\hat{\mathbf{W}} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{H}}'\hat{\mathbf{H}} & -\hat{\mathbf{H}}'\hat{\mathbf{W}} \\ -\hat{\mathbf{W}}'\hat{\mathbf{H}} & (\mathbf{I} + \lambda^2\mathbf{D}'\mathbf{D})\hat{\mathbf{W}}'\hat{\mathbf{W}} \end{bmatrix}. \tag{10}$$

The model and the data vectors are $\mathbf{m} = \begin{bmatrix} \hat{\mathbf{r}} & \mathbf{\Delta b} \end{bmatrix}^T$ and $\mathbf{d} = \begin{bmatrix} \hat{\mathbf{r}}_m^0 & -\lambda\mathbf{D}\hat{\mathbf{r}}_m^0 \end{bmatrix}^T$ respectively. Note that we no longer have restrictions about the λ value that would induce ill-posedness (other than make it excessively large). This scheme sucessfully passes the dot product test, which did not occur with non-extended LWIVU. However, now there arise three challenges:

1. Higher computational cost and storage because of extending: This problem becomes exacerbated in the 3-D case.
2. Unstable solution: DSO needs careful preconditioning to avoid artifacts, particularly in a linear problem.
3. Stacking extended volumes: Conventional stacking software dramatically slows down when extending.

In the next section I elaborate about these challenges.

EXTENDED LWIVU

Computational cost of subsurface offset extension

Extending the migration image domain in subsurface offset (hence producing offset-domain common image gathers, ODCIG) offers attractive features over the non-extended approach, such as going further and computing angle-domain common image gathers (ADCIG) for amplitude versus offset/azimuth analyses, and enabling migration velocity analysis. Meanwhile, time-lag extension has been proposed as one solution for the slow convergence of least-squares migration in the presence of velocity inaccuracies (**?**). In this report we have seen that it becomes possible to recast LWIVU in such a way that the occurence of ill-posedness as the result of minimizing a functional whilst maximazing the other is avoided altogether.

However, expanding in subsurface offset adds computational cost that is proportional to the total number of subsurface offset samples. Expansion operates exclusively during the scattering and imaging condition stages. Yet, it can easily overcome the computation time that it takes for wavefields propagations. For such a reason, I implement the so-called Nyquist approach (Sun and Fu, 2013). It consists of applying scattering or imaging only at time intervals fine enough to honor the maximum frequency in the data. For example, if the Courant-Friedrichs-Lewy (CFL) condition dictates that the wave propagation time step must be kept no greater than 1 ms, but the maximum frequency in the data is 40 Hz, then scattering/imaging can be applied every 12.5 ms, i.e., about once every twelve CFL time steps. Therefore, the impact of these costly tasks shrinks to one twelveth of the conventional approach. While this scheme works well for conventional RTM imaging, numerical experiments suggest that for seismic inversion it is safer to remain somewhat above the higher frequency. In the previous example this means that one may want to scatter/image every 7-9 CFL time steps. The reason for this, more conservative, approach is that decimating the scatter/imaging time step can produce numerical errors that affect the accuracy of the dot-product test and, as a consequence, the inversion result.

Differential semblance optimization in linear velocity inversion

Since its conception (Symes and Carazzone, 1991) DSO has been, to my knowledge, almost exculsively applied to non-linear velocity inversion (e.g., Vyas and Tang, 2010; Tang, 2011; Lameloise et al., 2015; Yang and Sava, 2015). DSO-based WEMVA typically consists of minimizing the following objective function:

$$\Phi(\mathbf{b}) = \frac{1}{2}\|\,|\mathbf{h}|\mathbf{r}_m(\mathbf{x},\mathbf{h};\mathbf{b})\|_{\mathbf{2}}^{\mathbf{2}}. \tag{11}$$

Here $\mathbf{b}$ represents the background model, containing both low-wavenumber and high-wavenumber components ($\mathbf{b}_0$ and $\mathbf{\Delta b}$, respectively), and $|\mathbf{h}|$ represents the size of the subsurface offset vector that constitutes here the DSO operator, $\mathbf{D}$. The corresponding gradient is

$$\nabla\Phi(\mathbf{b}) = \left[\frac{\mathrm{d}\mathbf{r}_m(\mathbf{b})}{\mathrm{d}\mathbf{b}}\right]' |\mathbf{h}|^2\mathbf{r}_m(\mathbf{b}). \tag{12}$$

I dropped the $\mathbf{x}$ and $\mathbf{h}$ dependencies to simplify the notation. It is important to notice that solving Equation 12 means that we invert for the *whole* background model $\mathbf{b}$. In contrast, for a linear problem we approximate the migrated image as

$$\mathbf{r}_m(\mathbf{b}) \approx \mathbf{r}_m(\mathbf{b}_0) + \left[\frac{\mathrm{d}\mathbf{r}_m(\mathbf{b})}{\mathrm{d}\mathbf{b}}\bigg|_{\mathbf{b}=\mathbf{b}_0}\right]\mathbf{\Delta b} = \mathbf{r}_m(\mathbf{b}_0) + \mathbf{W}\mathbf{\Delta b} \tag{13}$$

and now we invert for *perturbations in the background* assuming that we know $\mathbf{b}_0$. Substituting Equation 13 in Equation 11 we obtain the linear WEMVA misfit function,

$$\Phi(\mathbf{\Delta b}) = \frac{1}{2}\|\mathbf{D}\big(\mathbf{W}\mathbf{\Delta b} + \mathbf{r}_m^0\big)\|_2^2, \tag{14}$$

and its gradient,

$$\nabla\Phi(\mathbf{\Delta b}) = \mathbf{W}'\mathbf{D}'\mathbf{D}\big(\mathbf{W}\mathbf{\Delta b} + \mathbf{r}_m^0\big). \tag{15}$$

I go back to using $\mathbf{D}$ to represent DSO because I shall experiment with functions other than $\mathbf{D} = |\mathbf{h}|$ in the future. For instance, I am testing tapers for high values of $|\mathbf{h}|$, in an effort reduce the influence of events not related to velocity inaccuracies. The latter are assume to be small (which keeps the method within the linear regime), therefore I do not expect to have significant contributions from large offset values.

It is a common practice to precondition DSO-based WEMVA with low-pass filters. I follow Tang (2011) and employ interpolation using B-splines to penalize high-wavenumber components when inverting for $\mathbf{\Delta b}$ (see Appendix). B-splines also can be used for non-extended WEMVA based on migration image power maximization. I also experiment this strategy for the WEMVA-DSO stage of extended LWIVU.

However, my implementation of linear DSO-based WEMVA still suffers from convergence issues. To illustrate the problem I chose a small section of the Sigsbee A model shown in Figure 1a. It includes a negative Gaussian velocity perturbation with minimum value of -400 ft/s as seen in Figure 1b, corresponding to the positive

slowness squared perturbation shown in Figure 1c. I synthesized data using such "true" velocity. For imaging purposes I employed a smoothed version shown in Figure 1d, which ignores the anomaly (it is assumed to be unknown). The result from conventional extended reverse-time migration (RTM) of the synthetic data is shown in zero-offset section in Figure 2a, and in an ODCIG at $x = 36000$ ft in Figure 2b. The latter is extracted at the center of the anomaly, and we observe unfocused energy around that depth. However, there is also low-wavenumber energy at the upper portion of the ODCIG. These artifacts are rejected by masks incorporated into the forward and adjoint operators.

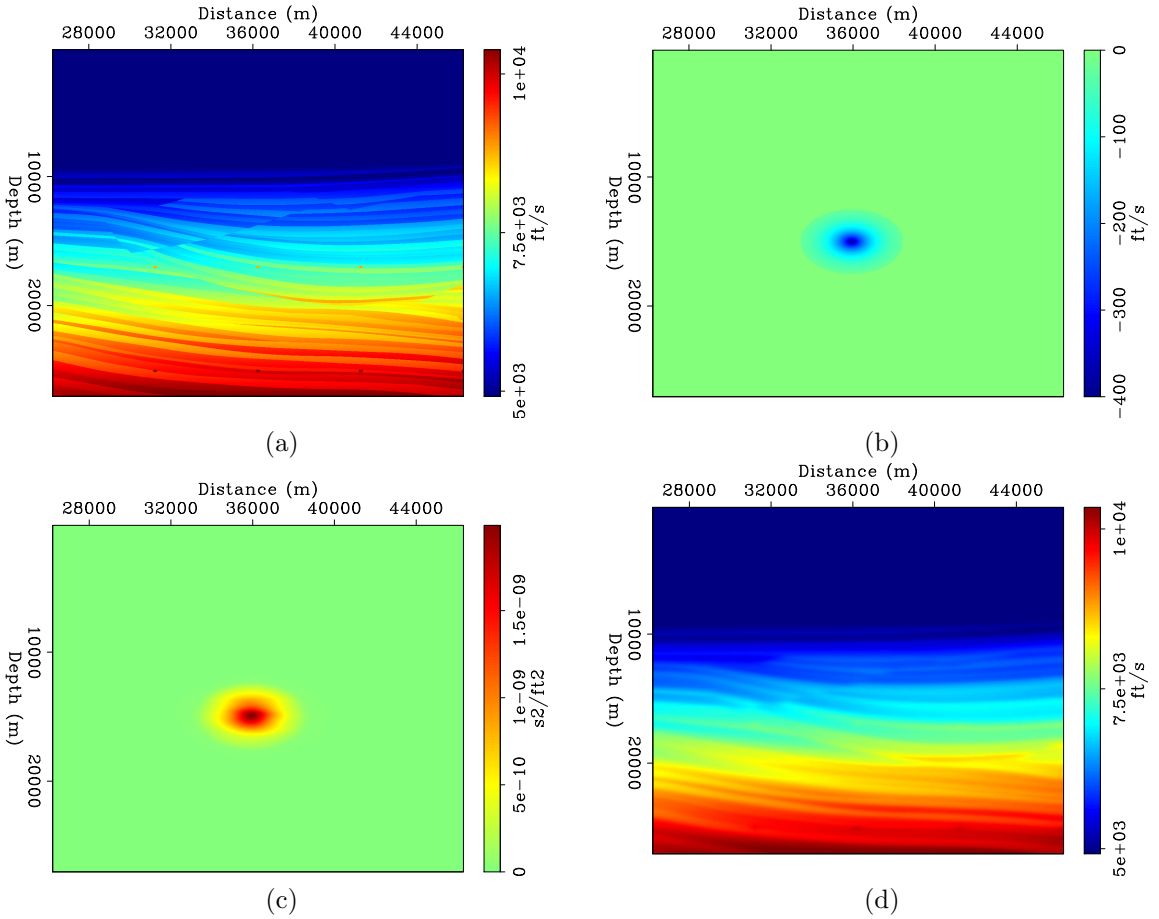

Figure 1:]
(a) Original velocity model extracted from the Sigsbee A model, including (b) a Gaussian negative anomaly in velocity. I synthesized data using this model. Hence, such data carry the information about the anomaly. (c) The same anomaly expressed in slowness squared becomes positive. (d) Smooth velocity employed for imaging, which does not include the velocity anomaly. [**ER**]

With the data and the migration image we can run the WEMVA inversion. I first performed non-extended WEMVA by maximizing the image power, using a non-extended RTM image (not shown here; it is virtually identical to Figure 2a). After 15 iterations we recover the perturbation in the background shown in Figure 3a, plotted

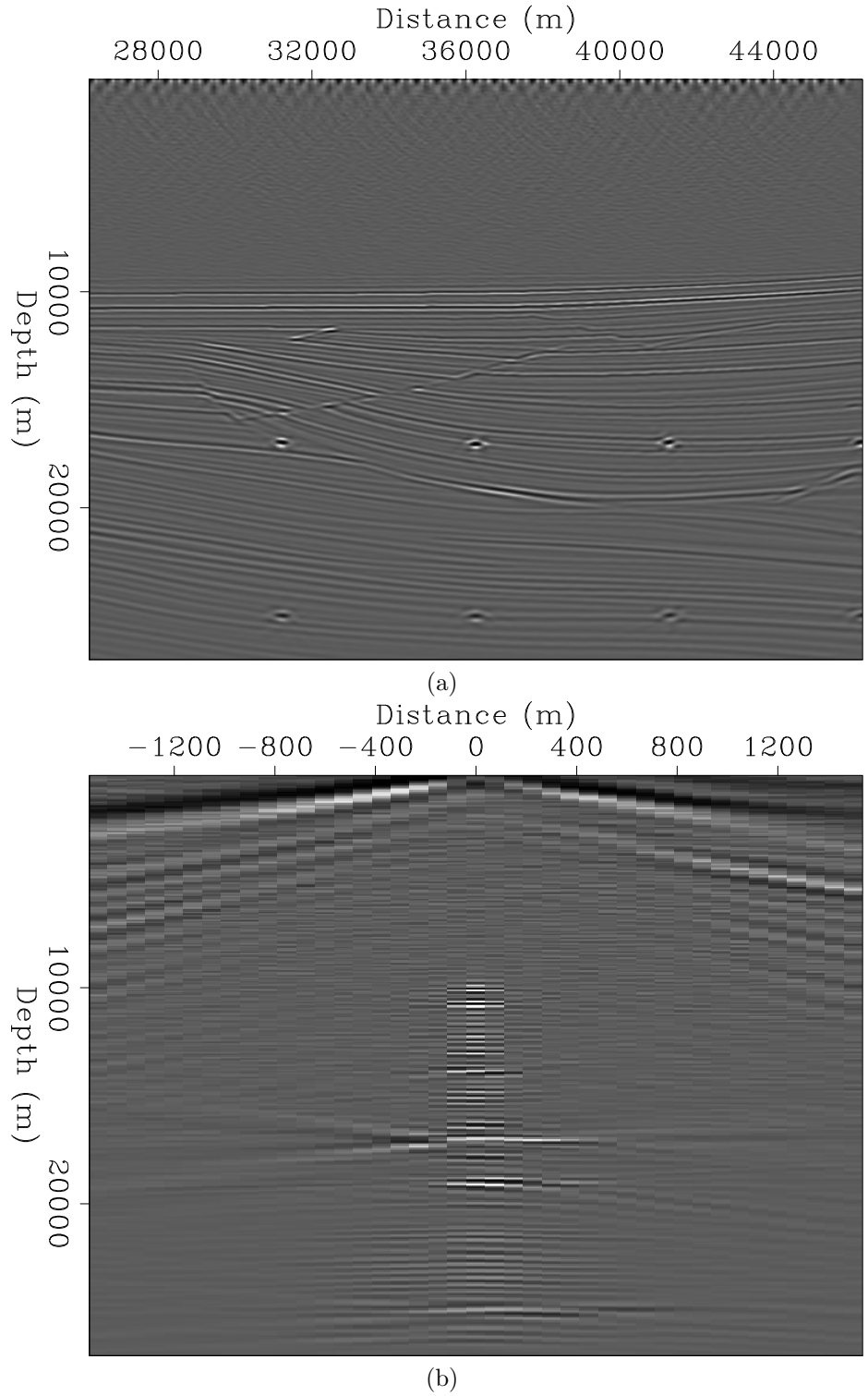

Figure 2:]
(a) Extended migration image of the Sigsbee A subset synthetic data, showing the zero subsurface offset section. (b) Same extended migration image, showing an ODCIG extracted at $x = 36000$ ft. The low-wavenumber noise at the upper part of the section is suppressed during the inversions by means of masks. [**CR**]

at the same scale as the true anomaly in slowness squared. We observe that the anomaly was satisfactorily recovered in amplitude and shape.

I also performed extended, DSO-based WEMVA using the extended RTM image (Figure 2a). After 15 iterations we recover the perturbation shown in Figure 3a, again plotted at the same scale as the true anomaly. Notice that this time I have failed in matching the amplitude of the anomaly, althought I indeed recovered its shape. Running the process for more iterations proved to be useless, as artifacts are introduced in the image. I am currently working on finding a solution to this problem.

Stacking extended volumes

The multiplicity of dimensions introduced by extending in subsurface offset slows down conventional stacking to the point that it takes longer than imaging. This issue is particularly notorious in 3-D, where two offset dimensions are employed. The inversion processes demand at least one stacking operation by iteration, therefore it represents an important bottleneck.

Bob Clapp (personal communication) suggested some solutions to this problem. I chose the use of MPI parallelization to distribute stacking among the processors. In a 3-D test that I performed, conventional staking took around 50 minutes to completion. For comparison, by using MPI stacking I was able to reduce the computational time to 8 minutes. Further time reduction is still possible.

FUTURE WORK

The current stumbling block is the inadequate perturbation in slowness squared that DSO-based WEMVA recovers. Tests that I will perform include the use of different DSO functions and possibly new masks to confine the updating areas of the gradient and the residual.

Another important issue is the estimation of the parameter λ. Claerbout (2014) and Biondi and Barnier (2017) have proposed different methods to estimate this parameter for balancing the fitting goals in a regularized problem of the type $\Phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{Lm} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{Rm}\|_2^2$. I will test both approaches for the LWIVU algorithm.

ACKNOWLEDGEMENT

Thanks to Petróleos Mexicanos for the financial support and to the SEP crew for fruitful comments and corrections.

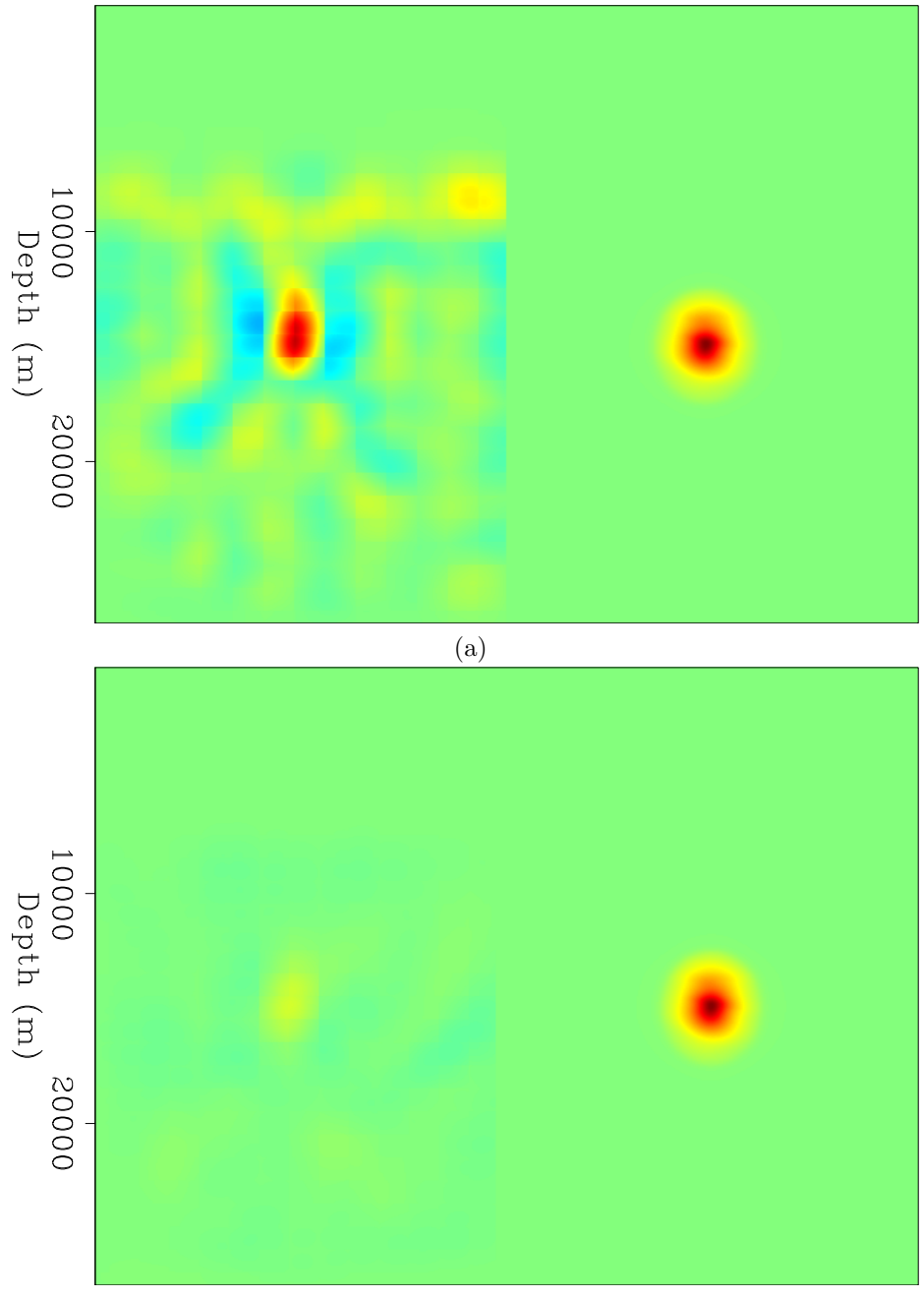

Figure 3:]
(a) Perturbation in the background recovered with non-extended WEMVA (left), in comparison with the true anomaly (right). (b) Perturbation in the background recovered with extended DSO-based WEMVA (left), in comparison with the true anomaly (right). In both figures, left and right panels were plotted at the same scale. [**CR**]

APPENDIX

Preconditioning LWIVU using B-splines

If we make the variable change

$$\mathbf{\Delta b} = \mathbf{B}\mathbf{\Delta p}, \tag{A-1}$$

where $\mathbf{\Delta p}$ is a coarse-grid representation of $\mathbf{\Delta b}$, and $\mathbf{B}$ represents interpolation using B-splines basis functions, then the DSO-based WEMVA misfit function becomes

$$\Phi(\mathbf{\Delta p}) = \frac{1}{2}\|\mathbf{D}\big(\mathbf{WB\Delta p} + \mathbf{r}_m^0\big)\|_2^2 \tag{A-2}$$

and the gradient becomes

$$\begin{aligned}\nabla\Phi(\mathbf{\Delta p}) &= \mathbf{B}'\hat{\mathbf{W}}'\mathbf{D}'\mathbf{D}\big(\hat{\mathbf{W}}\mathbf{B\Delta p} + \hat{\mathbf{r}}_m^0\big)\\ &= \mathbf{B}'\hat{\mathbf{W}}'\mathbf{D}'\mathbf{D}\hat{\mathbf{W}}\mathbf{B\Delta p} + \mathbf{B}'\hat{\mathbf{W}}'\mathbf{D}'\mathbf{D}\hat{\mathbf{r}}_m^0\\ &= \mathbf{B}'\underset{\sim}{\hat{\mathbf{W}}}'\underset{\sim}{\hat{\mathbf{W}}}\mathbf{B\Delta p} - \mathbf{B}'\underset{\sim}{\hat{\mathbf{W}}}'\underset{\sim}{\hat{\mathbf{r}}}_m^0,\end{aligned} \tag{A-3}$$

where $\underset{\sim}{\hat{\mathbf{W}}} = \mathbf{D}\hat{\mathbf{W}}$, $\underset{\sim}{\hat{\mathbf{W}}}' = \hat{\mathbf{W}}'\mathbf{D}'$, and $\underset{\sim}{\hat{\mathbf{r}}}_m^0 = -\mathbf{D}\hat{\mathbf{r}}_m^0$. Notice that Equation A-3 has the normal equations structure, including the minus sign. The data are the negative of the extended migration image after DSO.

We can proceed in a similar fashion in LWIVU. However, this time the DSO operator is not applied everytime we compute WEMVA. We start with the preconditioned version of Equation 7

$$\Phi(\hat{\mathbf{r}}, \mathbf{\Delta p}) = \frac{1}{2}\|\hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{B\Delta p} - \hat{\mathbf{r}}_m^0\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{D}\big(\hat{\mathbf{W}}\mathbf{B\Delta p} + \hat{\mathbf{r}}_m^0\big)\|_2^2. \tag{A-4}$$

The corresponding gradient is given by

$$\begin{aligned}\nabla\Phi = \begin{bmatrix}\nabla_{\hat{\mathbf{r}}}\Phi\\ \nabla_{\mathbf{\Delta p}}\Phi\end{bmatrix} &= \begin{bmatrix}\hat{\mathbf{H}}'\big(\hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{B\Delta p} - \hat{\mathbf{r}}_m^0\big)\\ -\hat{\mathbf{B}}'\hat{\mathbf{W}}'\big(\hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{B\Delta p} - \hat{\mathbf{r}}_m^0\big) + \lambda^2\hat{\mathbf{B}}'\hat{\mathbf{W}}'\mathbf{D}'\mathbf{D}\big(\hat{\mathbf{W}}\mathbf{B\Delta p} + \hat{\mathbf{r}}_m^0\big)\end{bmatrix}\\ &= \begin{bmatrix}\hat{\mathbf{H}}' & \mathbf{0}\\ -\hat{\mathbf{B}}'\hat{\mathbf{W}}' & \lambda\hat{\mathbf{B}}'\hat{\mathbf{W}}'\mathbf{D}'\end{bmatrix}\begin{bmatrix}\hat{\mathbf{H}}\hat{\mathbf{r}} - \hat{\mathbf{W}}\mathbf{B\Delta p} - \hat{\mathbf{r}}_m^0\\ \lambda\mathbf{D}\big(\hat{\mathbf{W}}\mathbf{B\Delta p} + \hat{\mathbf{r}}_m^0\big)\end{bmatrix}\\ &= \begin{bmatrix}\hat{\mathbf{H}}' & \mathbf{0}\\ -\hat{\mathbf{B}}'\hat{\mathbf{W}}' & \lambda\hat{\mathbf{B}}'\hat{\mathbf{W}}'\mathbf{D}'\end{bmatrix}\left\{\begin{bmatrix}\hat{\mathbf{H}} & -\hat{\mathbf{W}}\mathbf{B}\\ \mathbf{0} & \lambda\mathbf{D}\hat{\mathbf{W}}\mathbf{B}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{r}}\\ \mathbf{\Delta p}\end{bmatrix} - \begin{bmatrix}\hat{\mathbf{r}}_m^0\\ -\lambda\mathbf{D}\hat{\mathbf{r}}_m^0\end{bmatrix}\right\}.\end{aligned} \tag{A-5}$$

The forward and adjoint operators and the model and the data vectors are defined as the non-preconditioned case.

REFERENCES

Biondi, E. and G. Barnier, 2017, A flexible out-of-core solver for linear/non-linear problems: SEP-Report, **168**, 295–308.

Cabrales-Vargas, A., 2018, Improving reflectivity using linearized waveform inversion with velocity updating: SEP-Report, **172**, 193–208.

Claerbout, J. F., 2014, Geophysical image estimation by example: Jon Claerbout.

Lameloise, C., H. Chauris, and M. Noble, 2015, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: Geophysical Prospecting, **63**, 391–404.

Sun, W. and L. Y. Fu, 2013, Two effective approaches to reduce data storage in reverse time migration: Computers & Geosciences, **56**, 69–75.

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

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

Vyas, M. and Y. Tang, 2010, Gradients for wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts, 4077–4081.

Yang, T. and P. Sava, 2015, Image-domain wavefield tomography with extended common-image-point gathers: Geophysical Prospecting, **63**, 1086–1096.