Theory

In this chapter, I present formal derivations of different implementations the JIRB algorithm. I first show the linearization from the most general case and the reason for its failure. Next, I present the original derivation from the FWI objective function. Next, I show the final nonlinear JIRB objective function. Finally, I present some implementation ideas in the extended domain.

DERIVATION OF JIRB: MOST GENERAL CASES

Let us recall equation (**??**) for conventional LWI in model space:

$$\Phi(\mathbf{r}) = \frac{1}{2}\left\|\mathbf{H}(\mathbf{b})\mathbf{r} - \mathbf{I}(\mathbf{b})\right\|_2^2. \tag{1}$$

The most general case of JIRB simply consist in the incorporation of $\mathbf{b}$ into the inversion:

$$\Phi(\mathbf{r},\mathbf{b}) = \frac{1}{2}\left\|\mathbf{H}(\mathbf{b})\mathbf{r} - \mathbf{I}(\mathbf{b})\right\|_2^2. \tag{2}$$

Note that both the Hessian and the migration image depend on the background model $\mathbf{b}$, which hereafter will correspond to the *accurate* background model that we do not know a priori.

In equation (2), the model space becomes expanded, and so it is the model null space (**?**). Therefore, it becomes necessary to incorporate another restriction to drive $\mathbf{b}$ away from spurious solutions. For zero subsurface offset imaging, we can maximize the energy of the migration image to constrain $\mathbf{b}$, obtaining

$$\Phi(\mathbf{r},\mathbf{b}) = \frac{1}{2}\left\|\mathbf{H}(\mathbf{b})\mathbf{r} - \mathbf{I}(\mathbf{b})\right\|_2^2 - \frac{\lambda}{2}\left\|\mathbf{I}(\mathbf{b})\right\|_2^2, \tag{3}$$

where λ is a trade-off parameter. The minus sign allows maximizing the second functional within a minimization scheme. The interpretation of equation (3) is that the reflectivity fits an updated migration image that gets improved as the background model is corrected. Now I substitute equation (**??**) into (3) to obtain

$$\Phi(\mathbf{r},\mathbf{b}_0+\mathbf{\Delta b}) = \frac{1}{2}\left\|\mathbf{H}(\mathbf{b}_0+\mathbf{\Delta b})\mathbf{r} - \mathbf{I}(\mathbf{b}_0+\mathbf{\Delta b})\right\|_2^2 - \frac{\lambda}{2}\left\|\mathbf{I}(\mathbf{b}_0+\mathbf{\Delta b})\right\|_2^2, \tag{4}$$

where hereafter $\mathbf{b}_0$ will represent the *inaccurate* background model, which we assume it is available. Now it is the perturbation $\mathbf{\Delta b}$ what becomes a model parameter in the inversion.

LINEARIZATION ATTEMPT

Equation (4) is nonlinear with respect to $\mathbf{\Delta b}$ in both the Gauss-Newton Hessian and the migration image. We can make use of the following expansions:

$$\mathbf{H}(\mathbf{b}_0 + \mathbf{\Delta b})\mathbf{r} = \mathbf{H}(\mathbf{b}_0)\mathbf{r} + O(\mathbf{\Delta b}, \mathbf{r}), \tag{5a}$$

$$\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b}) = \mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} + O(\|\mathbf{\Delta b}\|^2). \tag{5b}$$

For a sufficiently small $\mathbf{\Delta b}$ we can drop the higher order terms, ($O(\mathbf{\Delta b}, \mathbf{r})$ and $O(\|\mathbf{\Delta b}\|)$). After doing that, equation (5a) means that the action of the Hessian upon the reflectivity when it is evaluated at the correct background, is not significantly different from the same action of the Hessian, but evaluated at the incorrect background. In this case the dropped higher order terms are second order in $\mathbf{r}$ and $\mathbf{\Delta b}$ simultaneously, which are both small with respect to $\mathbf{b}_0$. To support this claim I performed numerical experiments with the sedimentary section of the Sigsbee 2A model. Figure 1 show the true velocity, a perturbation in velocity, and the true reflectity. Figures 2, 3, and 4 compare the terms of expression (5a) for Gaussian anomalies similar to that of Figure 1b, but with peak values of -200, -800, and -1600 ft/s for the Gaussian anomaly. Notice that the difference becomes evident only at 1600 ft/s. This value represents an error of approximately 20% with respect to the background average, which is larger than the residual anomalies intended for the either the WEMVA or the JIRB methods.

Expression (5b) constitutes the Taylor's series expansion of the migration image around the wrong background model, which we can truncate after the first-order term in $\mathbf{\Delta b}$. $\mathbf{W}$ represents the forward WEMVA operator, which is the derivative of the migration image with respect to the background model. This expression is valid for perturbation values such that $\|\mathbf{\Delta b}\| \ll \|\mathbf{b}_0\|$.

Substituting equations (5a) and (5b) into equation (4) yields

$$\Phi(\mathbf{r}, \mathbf{\Delta b}) = \frac{1}{2}\left\|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{I}(\mathbf{b}_0) - \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}\right\|_2^2 - \frac{\lambda}{2}\left\|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}\right\|_2^2. \tag{6}$$

The first functional in equation (6) has the expected behavior of minimizing the difference between the reflectivity under the action of the Hessian, and the migration image evaluated at the wrong background, corrected with the WEMVA term applied to the perturbation in the background. This term becomes a perturbation in the image that contains the corrections to the background model.

However, the second functional in equation (6) does not have the expected behavior. The reason is because, whereas the linearization in expression (5b) is valid for small $\mathbf{\Delta b}$ values, substituting such a linearized term into the norm gives

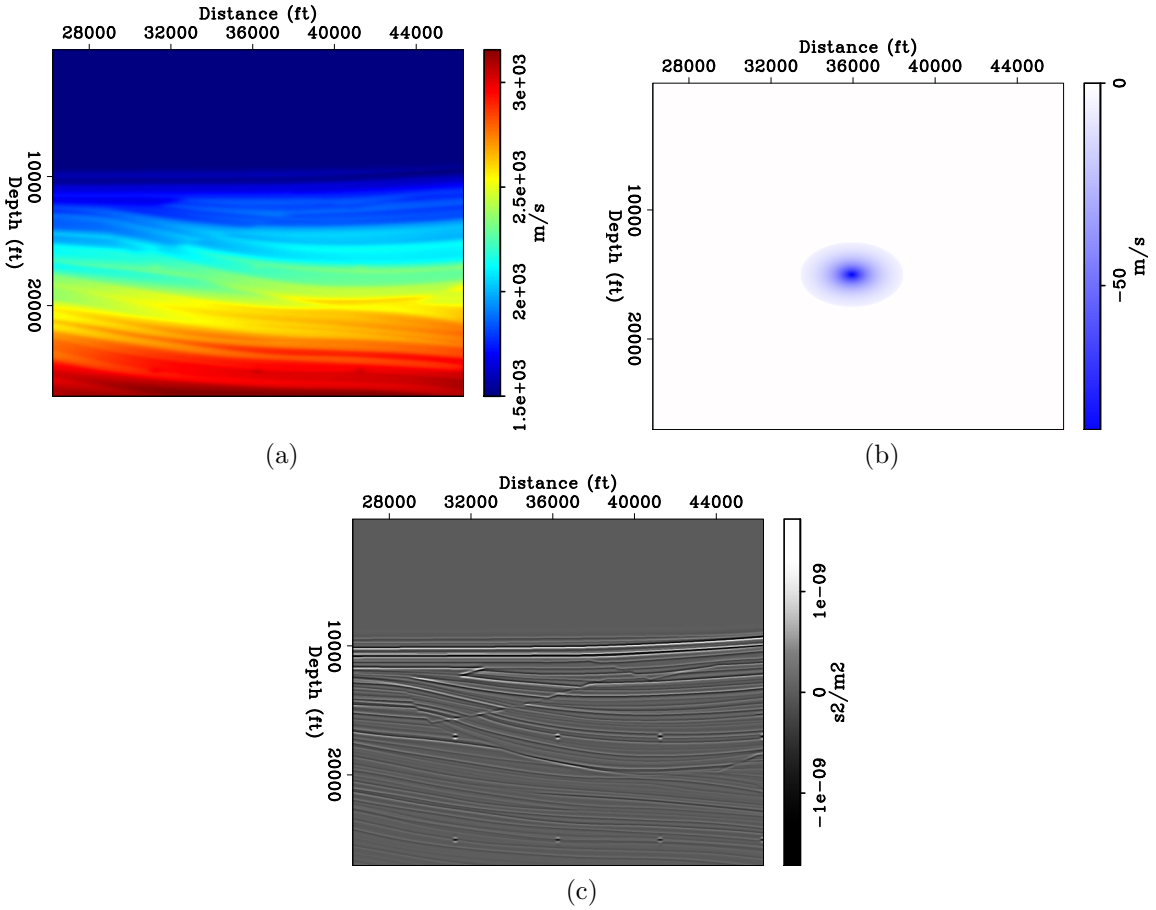


Figure 1: Sedimentary section of the Sigsbee model, showing: a) True velocity model. b) Gaussian anomaly. c) True reflectivity. [**CR**]

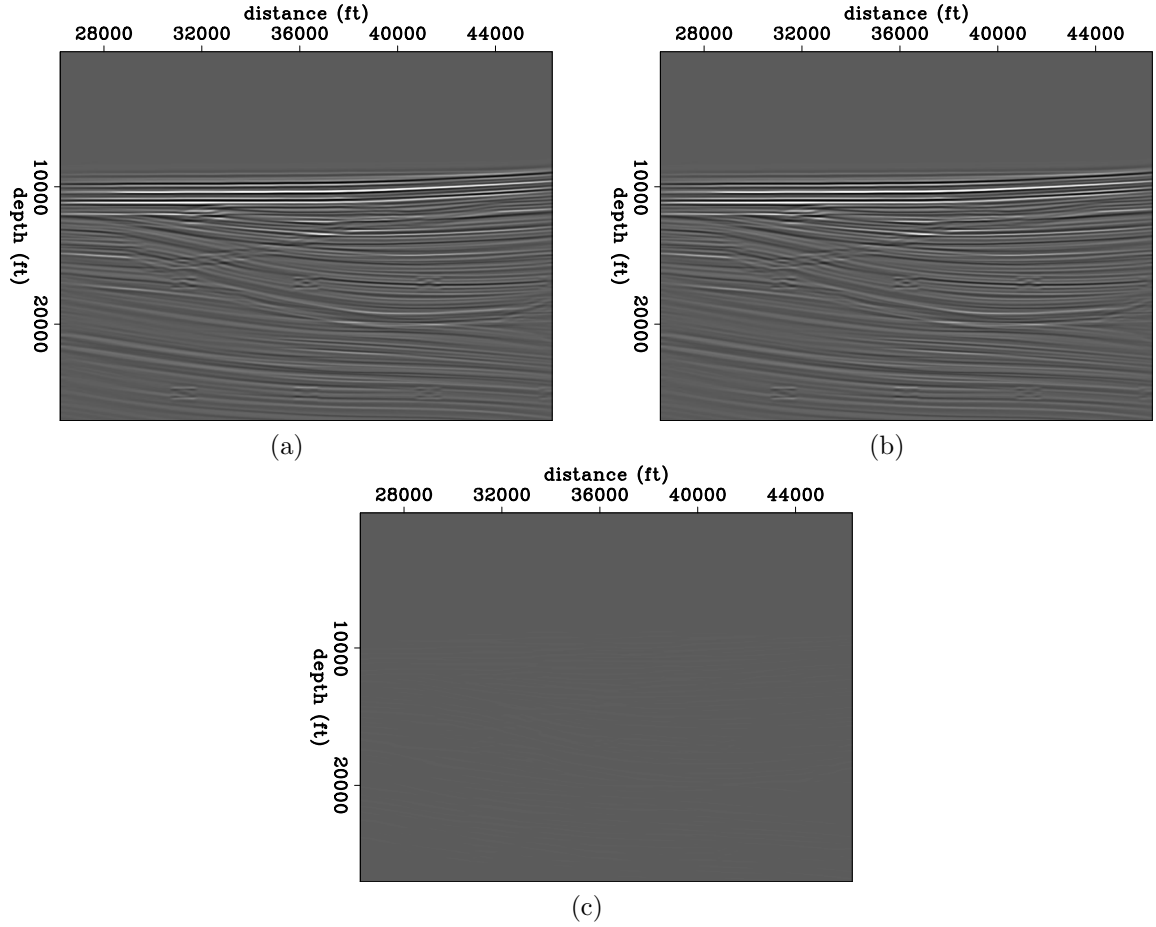


Figure 2: Test on the validity of expression (5a) for the Sigsbee model sedimentary section, using a Gaussian velocity perturbation of 200 ft/s: a) $\mathbf{H}(\mathbf{b})\mathbf{r}$; b) $\mathbf{H}(\mathbf{b}_0)\mathbf{r}$. c) Difference between images a) and b). [**CR**]

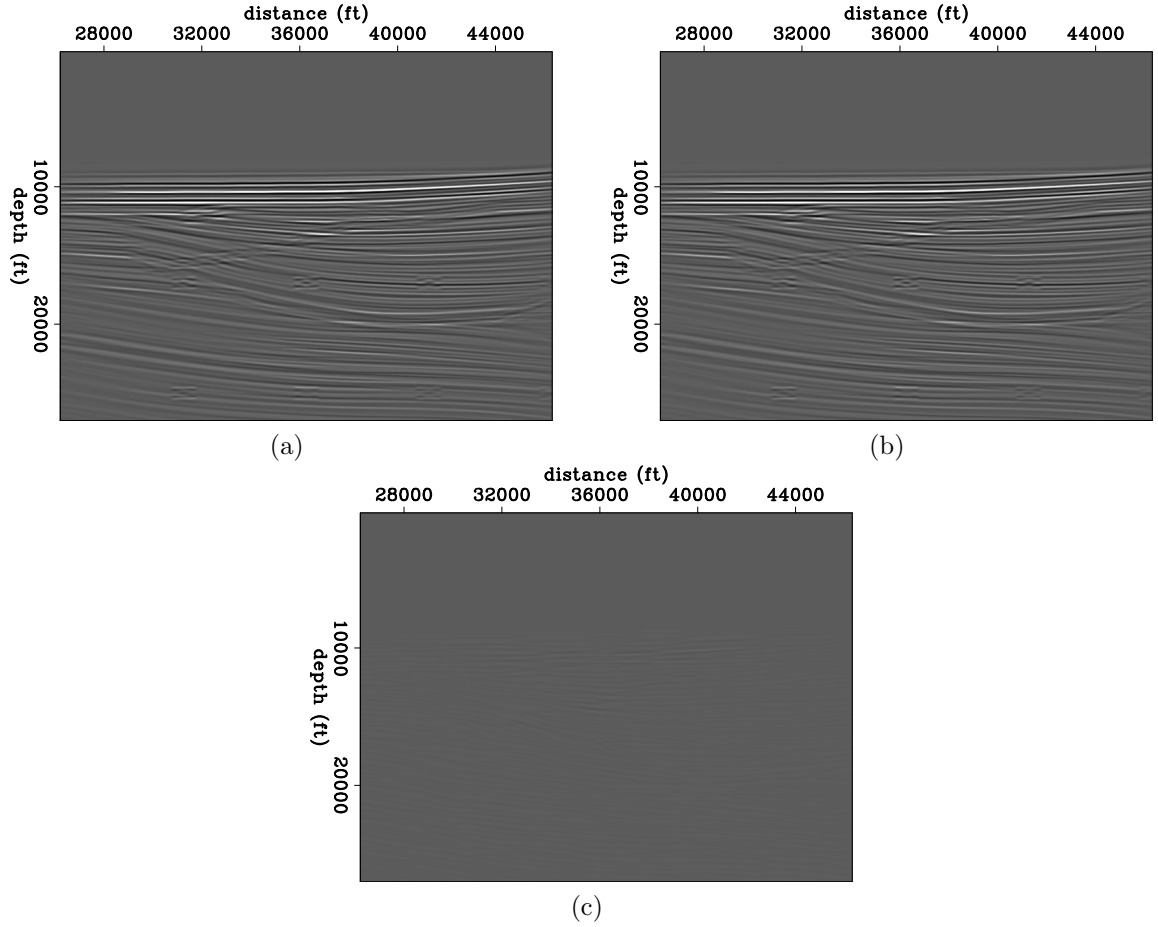


Figure 3: Test on the validity of expression (5a) for the Sigsbee model sedimentary section, using a Gaussian velocity perturbation of 800 ft/s: a) $\mathbf{H}(\mathbf{b})\mathbf{r}$; b) $\mathbf{H}(\mathbf{b}_0)\mathbf{r}$. c) Difference between images a) and b). [**CR**]

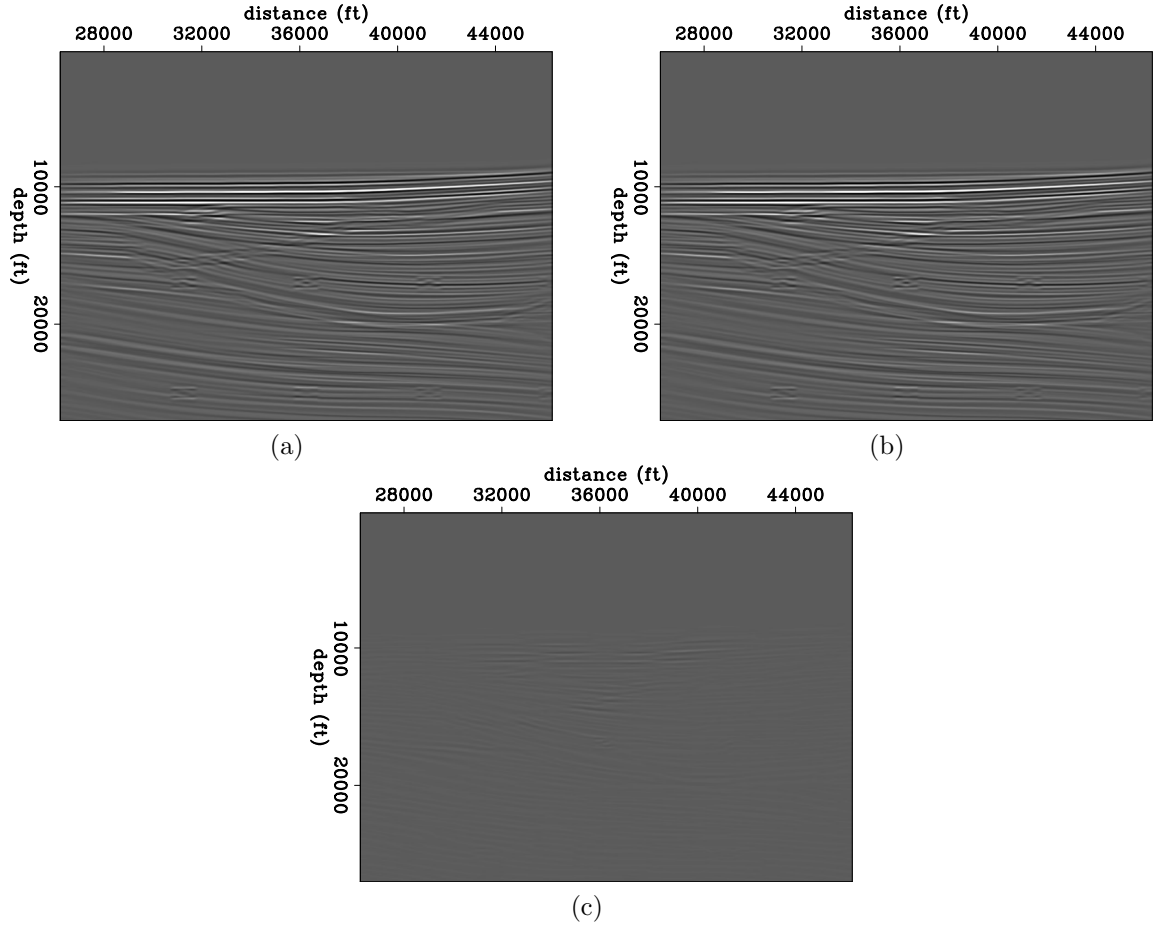


Figure 4: Test on the validity of expression (5a) for the Sigsbee model sedimentary section, using a Gaussian velocity perturbation of 1600 ft/s: a) $\mathbf{H}(\mathbf{b})\mathbf{r}$; b) $\mathbf{H}(\mathbf{b}_0)\mathbf{r}$. c) Difference between images a) and b). [**ER**]

$$\|\mathbf{I}(\mathbf{b}_0+\mathbf{\Delta b})\|_2^2 \approx \|\mathbf{I}(\mathbf{b}_0)+\mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b}\|_2^2\,, \tag{7}$$

which is inaccurate. In fact, expression (7) preserves a second order term in $\mathbf{\Delta b}$, but neglects another (see Appendix **??** for details). The consequence is that the linearized norm in the right-hand side of expression (7) behaves as a quadratic function of the type Ax^2+Bx+C, with curvature term $A \geq 0$ because it is represented by a positive semidefinite matrix (**??**). In other words, the linearized norm is *convex*, thus there is no maximum at the true background, as it is expected from the original functional (left-hand side of expression (7)), which is *concave*.

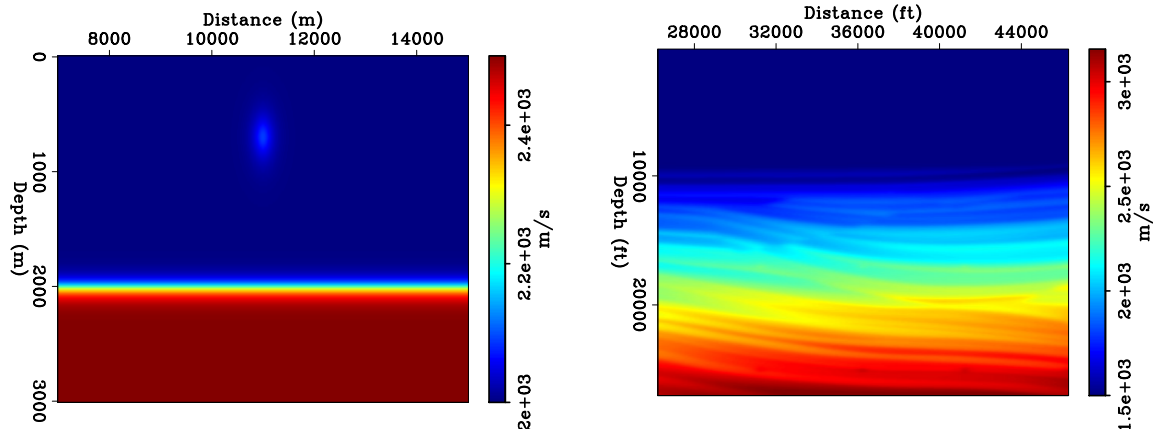
To further illustrate the problem, I performed numerical tests on a flat-layer model and the sedimentary sector of the Sigsbee model (Figure 5). The flat-layer model includes a positive Gaussian anomaly in velocity. This model represents the true background model, $\mathbf{b}^{\text{true}} = \mathbf{b}_0 + \mathbf{\Delta b}^{\text{true}}$. 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 does not contain a Gaussian anomaly. However, to construct an incorrect background model, I added a negative Gaussian anomaly in velocity to the true background model. The Gaussian anomalies for both models are shown in Figure 7.

I synthesized true data $\mathbf{d}^{\text{true}}$ in both cases by applying Born modeling to the corresponding reflectivity models (Figure 8) 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.

The first test demonstrates that the image power maximization term in the linearized objective function (right-hand side of expression (5b)) does not have the desired property of concavity. To do this, I created 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 $-1, -0.9, ..., 0.9, 1$. Then I evaluated the functional, $\big\|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\big[\mathbf{\Delta b}^{\text{true}} + \alpha\mathbf{\Delta b}_{\text{rand}}\big]\big\|_2^2$, 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.

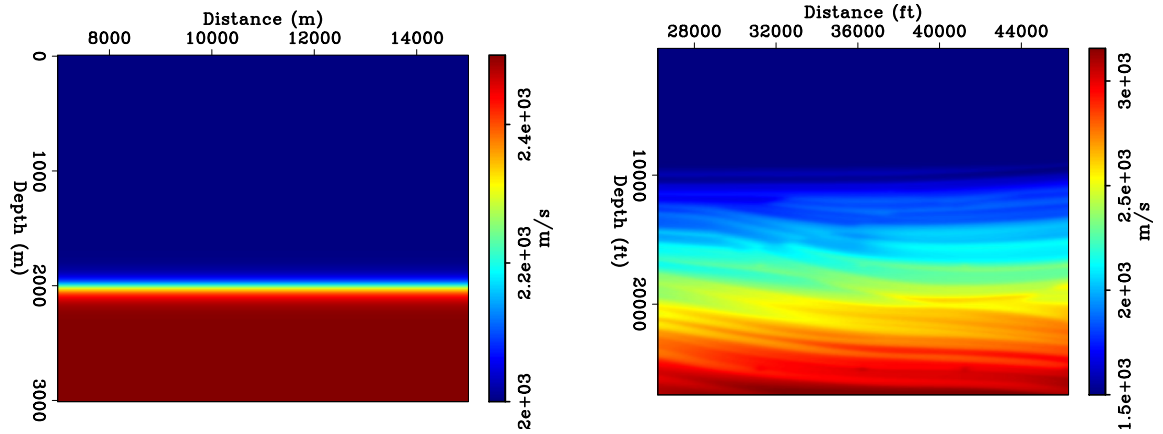
Figure 9 shows the result of 50 tests of the evaluation of the objective function using random directions, $\mathbf{\Delta b}_{\text{rand}}$, each one for the 21 values of α. Notice that there is no upper bound, and although the minima of the curves cluster around the vicinity of the true anomaly, none of them corresponds to it. These results prove that this objective function is convex, as I discussed before.

I performed similar tests for the fully nonlinear objective function of image power maximization (left-hand side of expression (5b)). Similar to the previous test, I evaluated this function at random directions. The results for both models are shown



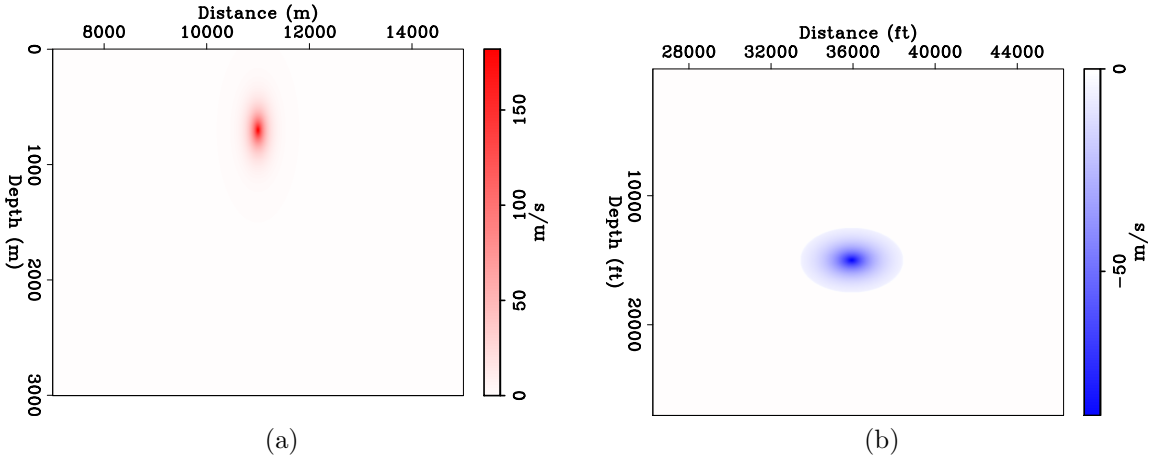
(a) (b)

Figure 5: True background models in velocity. a) Flat model. b) Sigsbee model. [**ER**]



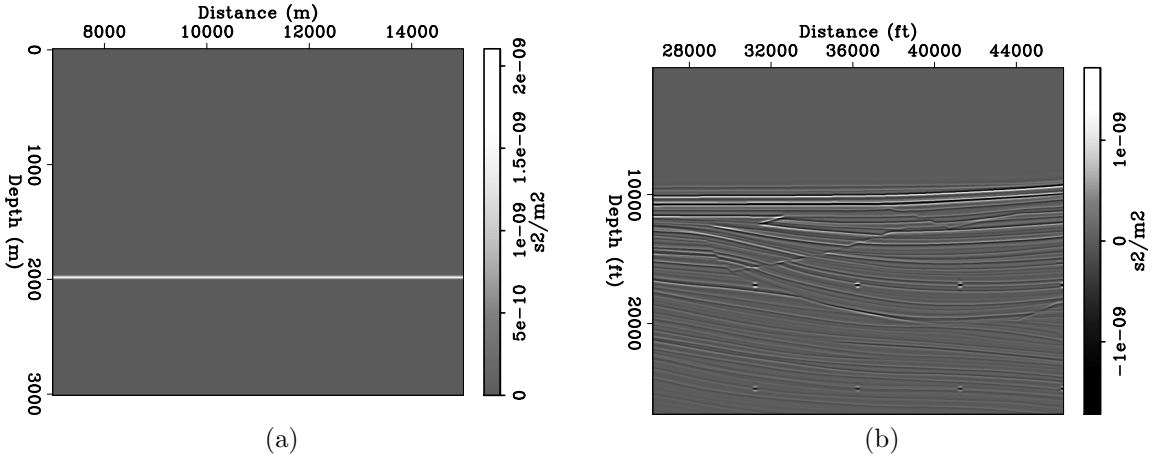
(a) (b)

Figure 6: Wrong background models in velocity. a) Flat model. b) Sigsbee model. [**ER**]



(a) (b)

Figure 7: Perturbations in the background models expresed in velocity. a) for the flat model. b) For the Sigsbee model. [**ER**]



(a) (b)

Figure 8: Reflectivity models in slowness squared. a) Flat model. b) Sigsbee model. [**ER**]

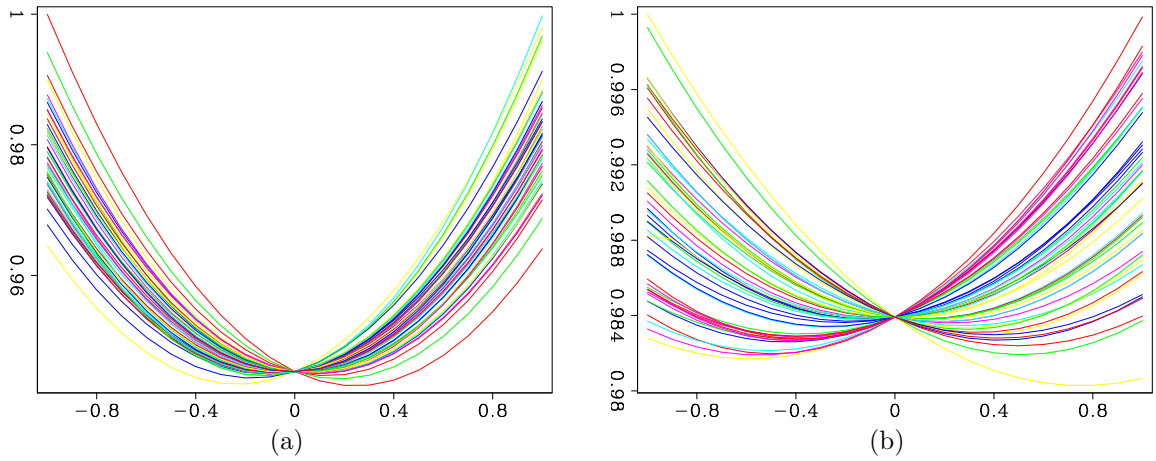


Figure 9: Experiment results of 50 random directions of the $\|\mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b}^{\text{true}} + \alpha\boldsymbol{\Delta}\mathbf{b}_{\text{rand}}\|_2^2$ functional in the a) flat model, and the b) Sigsbee model. Each random direction was explored with the α scalar. [**NR**]

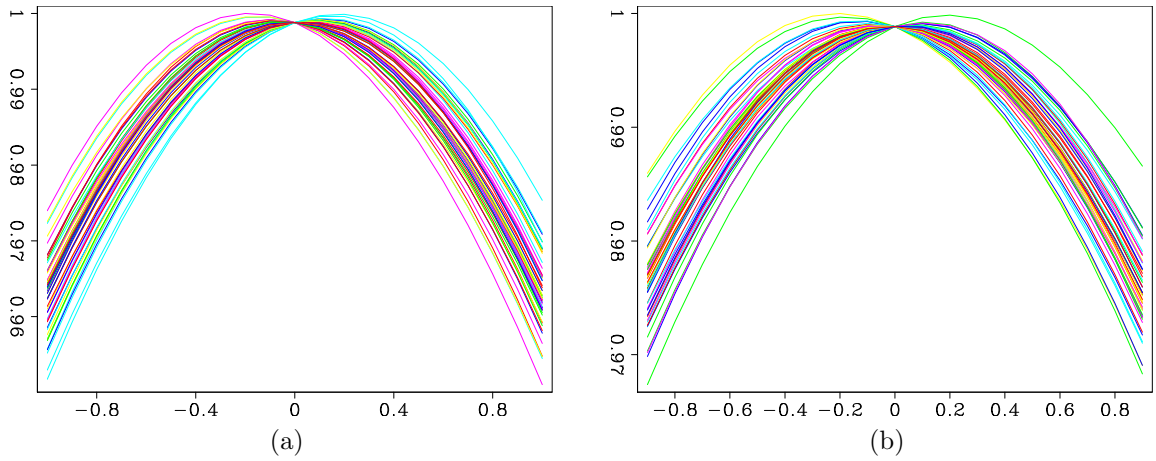


Figure 10: Experiment results of 50 random directions of the $\|\mathbf{I}(\mathbf{b}^{\text{true}} + \alpha\boldsymbol{\Delta}\mathbf{b}_{\text{rand}})\|_2^2$ functional in the a) flat model, and the b) Sigsbee model. Each random direction was explored with the α scalar. [**NR**]

in Figure 10. These results confirm that this objective function is concave along the directions explored. While these experiments do not tell us for sure if it is, in fact, concave, we can still conclude that this may be an approximate heuristic. However, even though the maximum values of the curves cluster around $\alpha = 0$, not all of them are exactly there.

DERIVATION FROM FULL-WAVEFORM INVERSION

The first attempt to envision a JIRB method was through the simplification of the Newton's equation associated to the FWI iterative solution (e.g. **??**).

Full-waveform inversion (**???**) is a nonlinear optimization technique that, in principle, can estimate all the wavenumber components of the subsurface model parameters, $\mathbf{m}$. However, in practice, it suffers from cycle skipping if the difference between the initial model and the actual model is more than half a wavelength in the data space. Multiscaling approaches are often employed to circumvent the local minima problem (e.g. **?**).

In a nutshell, FWI aims to minimize in the least-squares sense the misfit between the observed seismic data and synthetic data corresponding to model $\mathbf{m}$, which we can express as

$$\Phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d}(\mathbf{m}) - \mathbf{d}_{\text{obs}}\|_2^2, \tag{8}$$

where

$$\mathbf{d}(\mathbf{m}) = \mathbf{S}\big[\mathcal{L}(\mathbf{m})\big] \tag{9}$$

is the modeled data calculated by using the full-wave propagation operator, $\mathcal{L}$, collected at the receivers' position by the sampling operator, $\mathbf{S}$.

The solution of the FWI optimization problem in equation (8) using gradient descent optimization schemes or Hessian based methods (**?**) requires the computation of the gradient of the misfit function, which is given by

$$\nabla_{\mathbf{m}}\Phi(\mathbf{m}) = \big[\nabla_{\mathbf{m}}\mathcal{L}(\mathbf{m})\big]^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\big[\mathbf{S}\mathcal{L}(\mathbf{m}) - \mathbf{d}_{\text{obs}}\big] = \mathbf{L}(\mathbf{m})^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\big[\mathbf{S}\mathcal{L}(\mathbf{m}) - \mathbf{d}_{\text{obs}}\big], \tag{10}$$

where $\nabla_{\mathbf{m}}$ constitutes the gradient with respect to the subsurface model parameters, $\mathbf{S}^{\mathrm{T}}$ is the adjoint operator of sampling (hence it injects data at the receivers' positions), and $\mathbf{S}\mathcal{L}(\mathbf{m}) - \mathbf{d}_{\text{obs}}$ represents the data residuals that are projected into the model space by means of the RTM operator, $\mathbf{L}(\mathbf{m})^{\mathrm{T}} = \big[\nabla_{\mathbf{m}}\mathcal{L}(\mathbf{m})\big]^{\mathrm{T}}$.

The FWI solution can be obtained by local linearization where the current subsurface model, $\mathbf{m}_{\text{i}}$, is updated by a direction vector $\delta\mathbf{m}$ such that $\|\delta\mathbf{m}\| << \|\mathbf{m}_{\text{i}}\|$, thereby exhibiting a local quadratic behavior. Now we can expand the gradient in Taylor's series around $\mathbf{m}_{\text{i}}$ and drop second- and higher-order terms, obtaining

$$\nabla_{\mathbf{m}}\Phi(\mathbf{m}_{\text{i}}+\delta\mathbf{m}) \approx \nabla_{\mathbf{m}}\Phi(\mathbf{m}_{\text{i}}) + \nabla_{\mathbf{m}}\nabla_{\mathbf{m}}\Phi(\mathbf{m}_{\text{i}})\delta\mathbf{m} = \nabla_{\mathbf{m}}\Phi(\mathbf{m}_{\text{i}}) + \mathbf{H}_{\text{fwi}}(\mathbf{m}_{\text{i}})\delta\mathbf{m}, \quad (11)$$

where $\mathbf{H}_{\text{fwi}}(\mathbf{m}_{\text{i}})$ is the FWI Hessian, not be confused with the FWI *Gauss-Newton* Hessian, $\mathbf{H} = \mathbf{L}^{\text{T}}\mathbf{L}$, introduced in a previous section. As mentioned before, solving equation (11) represents the optimization of a quadratic problem in $\delta\mathbf{m}$. Therefore, the gradient evaluated at the updated model, $\mathbf{m}_{\text{i+1}} = \mathbf{m}_{\text{i}} + \delta\mathbf{m}$, becomes zero, and equation (11) reduces to

$$\mathbf{H}_{\text{fwi}}(\mathbf{m}_{\text{i}})\delta\mathbf{m} = -\nabla_{\mathbf{m}}\Phi(\mathbf{m}_{\text{i}}). \quad (12)$$

Equation (12) represents the well-known Newton's equation (**?**).

Now I will incorporate equation (**??**) into Newton's equation by defining the current model and the model step as follows:

$$\begin{aligned} \mathbf{m}_{\text{i}} &= \mathbf{b}_0, \\ \delta\mathbf{m} &= \boldsymbol{\Delta}\mathbf{b} + \mathbf{r}. \end{aligned} \quad (13)$$

On the other hand, according to **?** we can split the FWI Hessian into the FWI Gauss-Newton Hessian and the adjoint of the WEMVA operator:

$$\mathbf{H}_{\text{fwi}} = \mathbf{H} + \mathbf{W}^{\text{T}}. \quad (14)$$

Substituting equations (13) and (14) into the Newton's equation yields

$$\mathbf{H}_{\text{fwi}}(\mathbf{m}_{\text{i}})\delta\mathbf{m} = \big[\mathbf{H}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\big](\boldsymbol{\Delta}\mathbf{b} + \mathbf{r}). \quad (15)$$

As far as I am assuming the remaining inaccuracies in the background model to be small, it is valid to employ maximization of the image energy. Therefore, I dropped the transpose symbol in $\mathbf{W}$ because the WEMVA operator is self-adjoint in the zero-subsurface offset domain.

Re-arranging terms in the right-hand side of equation (15) we obtain

$$\big[\mathbf{H}(\mathbf{b}_0)\mathbf{r} + \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b}\big] + \big[\mathbf{H}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b} + \mathbf{W}(\mathbf{b}_0)\mathbf{r})\big]. \quad (16)$$

The first part inside brackets in expression (16) contains the image term $\mathbf{H}(\mathbf{b}_0)\mathbf{r}$ employed in model-space LWI, and the term $\mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b}$ that represents the perturbation in the image regarding the correction of the background model. I keep both terms because they have a predictable impact on primary events related to the subsurface reflectivity. The second part in brackets very likely contributes to second-order scattering events $\big[\mathbf{W}(\mathbf{b}_0)\mathbf{r}\big]$ and refinement of the low-wavenumber component $\big[\mathbf{H}\boldsymbol{\Delta}\mathbf{b}\big]$. I drop the first term because it is not related to the subsurface reflectivity's primary events, and I drop the second one because the method does not aim to retrieve a high-wavenumber background model.

Regarding the FWI gradient in the Newton's equation (right-hand side of equation (12)), it becomes $\nabla\Phi(\mathbf{b}_0)$. Substituting this gradient in equation (10) yields

$$\nabla\Phi(\mathbf{b}_0) = \big[\nabla\mathcal{L}(\mathbf{b}_0)\big]^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\big[\mathbf{S}\mathcal{L}(\mathbf{b}_0) - \mathbf{d}_{\mathrm{obs}}\big]. \tag{17}$$

Nonlinear wave propagation at the background model (free of reflections) will model only direct waves and diving waves. Therefore, the term $\mathbf{S}^{\mathrm{T}}\big[\mathbf{S}\mathcal{L}(\mathbf{b}_0) - \mathbf{d}_{\mathrm{obs}}\big]$ simply represents the negative of the perturbation in the data, $\boldsymbol{\Delta}\mathbf{d}_{\mathrm{obs}}$, (**?**), i.e., the data after removing the direct and diving waves. On the other hand, the adjoint of the gradient term represents Born adjoint modeling or RTM evaluated at $\mathbf{b}_0$. Thus, equation (17) becomes the negative of the RTM image,

$$\nabla\Phi(\mathbf{b}_0) = -\mathbf{L}(\mathbf{b}_0)^{\mathrm{T}}\boldsymbol{\Delta}\mathbf{d}_{\mathrm{obs}} = -\mathbf{I}(\mathbf{b}_0). \tag{18}$$

Substituting the first term in braces in expression (16) and equation (18) in the Newton's equation we obtain

$$\mathbf{H}(\mathbf{b}_0)\mathbf{r} + \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b} = \mathbf{I}(\mathbf{b}_0), \tag{19}$$

which I named Linearized Waveform Inversion with Velocity Updating (LWIVU), and is equivalent to minimizing the objective function

$$\Phi(\mathbf{r}, \boldsymbol{\Delta}\mathbf{b}) = \frac{1}{2}\left\|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{I}(\mathbf{b}_0) + \mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b}\right\|_2^2. \tag{20}$$

Note that this objective function represents the first functional in equation (6). The only difference is in the positive sign for $\mathbf{W}(\mathbf{b}_0)\boldsymbol{\Delta}\mathbf{b}$, which is not a problem because the perturbation in the background can take positive or negative values.

NONLINEAR SCHEME

Given the shortcoming of the objective function (6), **?** proposed to keep the nonlinear objective function for the migration image, still maintaining 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. Therefore, the JIRB objective function became

$$\Phi(\mathbf{r}, \mathbf{\Delta b}) = \|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{W}(\mathbf{b}_0)\mathbf{\Delta b} - \mathbf{I}(\mathbf{b}_0)\|_2^2 - \lambda \|\mathbf{I}(\mathbf{b}_0 + \mathbf{\Delta b})\|_2^2 . \tag{21}$$

This change 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.

Although the objective function (21) behaves as desired, it is more straightforward and sensible to implement the non-linearity in both objective functions:

$$\Phi(\mathbf{r}, \mathbf{b}) = \|\mathbf{H}(\mathbf{b}_0)\mathbf{r} - \mathbf{I}(\mathbf{b})\|_2^2 - \lambda \|\mathbf{I}(\mathbf{b})\|_2^2 . \tag{22}$$

Objective function (22) can be solved (for local minima) using gradient-based methods such as steepest descent, nonlinear conjugate gradients, or quasi-Newton (**?**). Notice that the Gauss-Newton Hessian is precomputed with the wrong background model, and does not require to be updated as it is the case for the migration image. Solving objective function (22) is the core of the JIRB method. Chapter 4 shows a variation where I merely add preconditioning and weighting factors.

EXTENDED DOMAIN

The joint inversion of reflectivity and perturbation in the background can also be cast in the extended domain as

$$\Phi(\tilde{\mathbf{r}}, \mathbf{\Delta b}) = \frac{1}{2} \left\| \tilde{\mathbf{H}}(\mathbf{b}_0 + \mathbf{\Delta b})\tilde{\mathbf{r}} - \tilde{\mathbf{I}}(\mathbf{b}_0 + \mathbf{\Delta b}) \right\|_2^2 + \frac{\lambda}{2} \left\| \mathbf{D}\tilde{\mathbf{I}}(\mathbf{b}_0 + \mathbf{\Delta b}) \right\|_2^2 , \tag{23}$$

where the tilde represents extension in subsurface offset and $\mathbf{D}$ represents the differential semblance optimization operator (DSO) (**?**). In this case, the second functional ought to be minimized to penalize non-zero subsurface offset energy originated from incorrect background model.

At first glance it appears that in the expanded case the second functional can be linearized by substituting $\mathbf{D}\tilde{\mathbf{I}}(\mathbf{b}_0 + \mathbf{\Delta b}) \approx \mathbf{D}\left[\tilde{\mathbf{I}}(\mathbf{b}_0) + \tilde{\mathbf{W}}(\mathbf{b}_0)\mathbf{\Delta b}\right]$ in the norm, because both the nonlinear and the linearized functionals are convex, in contrast to the non-extended case where we maximize the image power, analyzed before. However, this approximation still lacks one second-order term that can lead to incorrect results (see Appendix **??**).

We can use a similar objective function as that of the zero subsurface offset domain:

$$\Phi(\tilde{\mathbf{r}}, \mathbf{b}) = \frac{1}{2} \left\| \tilde{\mathbf{H}}(\mathbf{b}_0)\tilde{\mathbf{r}} - \tilde{\mathbf{I}}(\mathbf{b}) \right\|_2^2 + \frac{\lambda}{2} \left\| \mathbf{D}\tilde{\mathbf{I}}(\mathbf{b}_0 + \mathbf{\Delta b}) \right\|_2^2 . \tag{24}$$

The obvious caveat of equation (24) is that inversion in the extended domain is more costly than the inversion in the zero offset domain because the scattering operator and the imaging condition operator dominate the computations. This problem is particularly severe in 3D imaging. Some people have proposed strategies to alleviate the problem, such as the so-called "Nyquist approach"(**?**), where we reduce to the minimum the number of times that the imaging condition is applied during RTM, depending on the frequency range of the data. However, in order to achieve enough extension, the computational burden is still considerable. The DSO operator often requires special preconditioning to avoid high-wavenumber artifacts in the gradient (e.g. **???**). For such reasons, in this dissertation I implement the JIRB methodology exclusively in the zero subsurface offset domain.