Full waveform inversion using wave-equation depth migration with tying to wells
Gary F. Margrave*, Robert J. Ferguson, and Chad M. Hogan
CREWES, University of Calgary, Calgary, Alberta

Summary
We examine the key concepts in full waveform inversion (FWI) and relate them to processes familiar to practicing geophysicists. For clarity, we present the central theoretical result behind FWI as a mathematical statement that a linear update to a migration velocity model is proportional to a prestack reverse-time migration of the data residual (the difference between the actual data and data predicted by the model) where the proportionality factor must be estimated. We argue that in many cases this factor will be complex-valued and frequency dependent, or in the time domain, it will be a convolutional wavelet. We find an analogy between estimation of the velocity update from the migrated section and the common process of impedance inversion, and we suggest that FWI can be viewed as a practical cycle of data modeling, migration of the data residual, and "calibration" of this migration to deduce the velocity update. The calibration step can be accomplished like a conventional impedance inversion where the migrated data residual is tied to the velocity residual (the difference between actual velocity and migration velocity) at a well. As there are a great many established algorithms for impedance inversion, so there are a plethora of possibilities for calibration. We present an extended example using the Marmousi model in which we use wave-equation migration (e.g. depth stepping) of the data residual and a simple least-squares amplitude scaling and constant phase rotation, determined at a simulated well, to calibrate the migration. We find that our approach produces a much improved velocity model in only a few iterations.

Introduction
Laïlly (1983) and Tarantola (1984) introduced full waveform inversion (FWI) to reflection seismology. As originally proposed, the procedure requires a sequence of pre-stack migrations where each migration is based on a velocity model updated by the result of the previous migration. More recently (e.g. Virieux and Operto 2009), FWI is formulated as a generalized inverse problem that requires only a forward modeling code, it's adjoint, and an iterative numerical-solver (usually a gradient method). While this approach is more general and efficient, lost is the perspective of FWI as a sequence of migrations and we return to that perspective here. We suggest that FWI can be viewed as an iterative cycle involving forward modeling, pre-stack migration, impedance inversion, and velocity model updating in each iteration.

Theory
The fundamental result that enables FWI was first derived by Laïlly (1983) and Tarantola (1984) and many others since then. Stated informally, the result says that, given a smooth migration velocity model, \( v_0(x,z) \), then a linearized update to the velocity model is given by

\[
\delta v_t(x,z) = \lambda \int \sum_{s,r} \varphi_t^*(x,z,\omega) \delta \varphi_{r(t),0}(x,z,\omega) d\omega
\]

where \( \lambda \) is a scalar constant or "step size", the hat (^) over a variable indicates its temporal Fourier transform, \( \varphi_t(x,z,\omega) \) is a model of the source wavefield for source \( s \) propagated to all \( (x,z) \), \( \delta \varphi_{r(t),0}(x,z,\omega) \) is the data residual for source \( s \) back propagated to all \( (x,z) \), \( * \) is complex conjugation, and \( \omega \) is temporal frequency. By data residual, we mean the difference between the actual data and synthetic data created by forward modeling through the velocity model \( v_0(x,z) \). The term on the right-hand side of equation 1 is, excluding the \( \lambda \) factor, a frequency-domain expression for a reverse-time migration in which both fields have been time-differentiated. In fact, equation 1 is often written as a time-domain expression given by

\[
\delta v_t(x,z) = \lambda \sum_{s,r} \int \frac{\partial \varphi_t^*(x,z,t)}{\partial T} \delta \varphi_{r(t),0}(x,z,T-t)dt.
\]

where \( T \) is the record length. If \( \lambda \) can be determined, then a new velocity model is estimated as

\[
v_t(x,z) = v_0(x,z) + \delta v_t(x,z).
\]

Given this basic result, the procedure is then iterated as shown in Figure 1. If the initial velocity model is sufficiently close to the true model, then convergence is possible and a solution to the nonlinear seismic inverse problem is obtained by a series of linearized steps.

The usual method of determining \( \lambda \) is called a "line search" (e.g. Pratt 1999). The logic is that the prestack migration, being the gradient of the misfit function, estimates the direction of descent on the error surface but the step size must be determined. The 1D search (line search) of a few trial step sizes and estimation of the optimal one determines \( \lambda \).
FWI using WEM with well tying

Fig. 1: The cycle of acoustic FWI (full waveform inversion). The cycle has external inputs at a) the initial velocity model, and b) the actual (or recorded) seismic data. The cycle counter $k$ is initially 1. Step 1) Velocity model $v_{k-1}$ is used to predict synthetic seismic data matching the acquisition geometry. Step 2) The data residual (real data - synthetic) is pre-stack migrated and stacked. Step 3) The pre-stack migration is "calibrated" to estimate a velocity perturbation $\delta v_k$. Step 4) The velocity model is updated by adding the perturbation to $v_{k-1}$ to estimate $v_k$.

The RTM part of equation 2 actually emerges in theory as the negative of the gradient of the data misfit function (the measure of the difference between the actual data and modelled data). In a gradient descent inversion iteration, the model is updated by stepping in the direction of the negative gradient (e.g. steepest descent) and $\lambda$ is the step size. The RTM in equation 2 is a conventional pre-stack RTM with two slight differences. First, the data residual is migrated rather than the data itself, and second, both the modelled shot field and the back propagated data residual have been time differentiated. This double time differentiation accounts for the factor for frequency squared in equation 1. The reason for the particular form is that, in theory, it makes the estimation of the velocity perturbation a simple scaling operation as in equation 1. However, this theory assumes the source waveform is known. If not, then $\lambda$ becomes frequency dependent, which in the time domain necessitates a convolution. Once $\lambda$ is frequency dependent, then the $\omega^2$ factor can be absorbed into it and we then restate equation 2 as

$$\delta v_k(x,z) = d(z) \cdot \int \psi_i(x,z,t) \delta \psi_{i(j,k)}(x,z,T-t) dt$$

(4)

where the RTM is now a standard algorithm (no time differentiation) and $d(z)$ is some kind of a deconvolution filter whose purpose is to convert the migrated section into a velocity estimate.

We have taken considerable liberties with equation 4 in assuming that $d(z)$ will only depend upon depth. In sufficiently simple media this is easily justified but in general it is more complex. We do this to draw a connection with the standard process of matching migrated data to wells in order to estimate impedance (Lindseth 1979). In a constant density setting, there is no difference between an impedance estimation and a velocity estimation so this leads to our conjectured generalization of FWI. Explicitly, we conjecture that the RTM in equation 4, and hence the migration algorithm of step 2 of Figure 1, can be replaced with any prestack depth migration and that $d(z)$ can be estimated by a suitable matching to well control and this constitutes step 3 of Figure 1. Matching migrated seismic sections to well control is a common step in interpretation, and it is intended to deal with residual wavelet amplitude and phase that remain after data processing. The intent of equation 4 is to convert the prestack migration of the data residual into a velocity perturbation. At a point of well control, the velocity residual, e.g. the difference between the true velocity and the migration velocity is known. So we propose to match the migrated data residual at the well to the velocity residual.

We have also found it useful to apply a spatial convolutional smoother to the migrated stack before matching to the well. The justification for this is to suppress migration artefacts that commonly arise in low-fold areas of the stack. This is commonly done in FWI and is known as "preconditioning the gradient".

Examples

To test our conjectures, using the Marmousi velocity model (Figure 2a), we created 40 shot records using finite-difference modeling (second order in time and space), with sources at the surface beginning at 4000 m, incrementing by 100 m, and extending to 7900 m. Each source record has a split-spread receiver pattern with offsets ranging from -2000 m to +2000 m at increments of 8.333 m. This constitutes our "real" data. As a migration algorithm, we used a pre-stack PSPI (phase-shift-plus-interpolation) algorithm (Gazdag and Squazerro 1984) implemented by us. PSPI is a space-frequency algorithm which means that each frequency is migrated independently. This has proven to be a very useful property because it facilitates an inversion process that is progressive in frequency by which we mean the low frequencies are inverted first and higher frequencies are inverted using the model as updated by the lower frequencies (as done by Pratt 1999). Finally, we assumed a well at coordinate 6000 m where the exact velocity is known over the depth interval [500, 2500] m.
Our calibration at the well is a simple two step process where we determine a constant scalar and a constant phase rotation by least squares to match the migrated trace at the well to the velocity residual.

In Figure 2 we show the results of a single iteration in which the migration was limited to the low frequency band 0-5 Hz. Figure 2b shows the migration velocity model used and Figure 2c is the updated model formed by adding the scaled and phase-rotated migrated stack to the migration model. The updated model will then become the migration model for the next iteration.

Figure 3 shows the calibration process at the well. The migrated data residual trace at the well has amplitudes between ±5x10⁻⁴ while the velocity residual (the subtraction of the curves in panel 3b) is 6 orders of magnitude greater. After scaling and phase rotation, the migrated trace is added to the migration model to produce a better approximation to the exact velocity (Figure 3c). In Figure 4 we show a selection of velocity models deduced at particular iterations in a test where the frequency band was progressively increased. The final velocity model is clearly an improvement on the initial one, but not fully correct indicating that convergence was not achieved.

Fig. 2: a) The Marmousi velocity model (black vertical line is the well). b) The initial migration velocity model formed by smoothing a) with a 600m Gaussian smoother. c) The updated migration velocity model after 1 iteration.

Fig. 3: The calibration process on the first iteration. a) The migrated data residual trace at the well. b) The exacted velocity (blue) and the migration velocity (red), c) The exact velocity (and the updated migration velocity at the well (red).

Fig. 4: A selection of velocity models as updated after a particular iteration of a 22 iteration test in which the frequency band was progressively increased from low to high.
FWI using WEM with well tying

Fig. 5: A sequence of velocity models resulting from an iteration in which the frequency band is kept constant (1-40Hz) but the size of the 2D Gaussian convolutional smoother applied to the migrated stack is varied. The smoother size is noted above each image. This is a study of the size of the convolutional smoother used to precondition the gradient.

Fig. 6. A plot of the L2 norm of the data residual versus iteration number for the experiment shown in Figure 4.

Figure 5 shows an alternative iteration in which the frequency band is kept fixed and the size of a 2D convolutional smoother used to precondition the gradient is varied. The smoother size is noted above each panel. The convolutional smoother is applied directly to the migrated stack before matching to well control and is useful to suppress migration artifacts. There is a dramatic improvement at iteration 7 which is where the smoother size drops below dominant wavelength of the data. Figure 6 shows a plot of the $L_2$ norm of the data residual versus iteration number for the experiment documented in Figure 5. The norm shows a sharp drop at iteration 7 and then continues to decrease until iteration 19. That it does not monotonically decrease is a result of our matching at the well instead of directly matching modelled data to raw data.

Conclusions

We have argued that each iteration in FWI (full waveform inversion) can be viewed as a four step process of (i) forward modeling through the migration velocity model, (ii) migration of the data residual (iii) calibration of the migration by converting it to a residual velocity estimate (iv) updating of the velocity model by adding the calibrated migration to it. We have conjectured that any prestack depth migration can be used and that the calibration, or conversion to velocity, can be done as an impedance inversion using well control. We have supported our conjectures with a study using synthetic data from the Marmousi model.

Acknowledgements

We thank the sponsors of CREWES for supporting this work. We also acknowledge the financial support of NSERC and MITACS.