A Perspective on Full-Waveform Inversion

Gary F. Margrave, Kris Innanen, and Matt Yedlin.

ABSTRACT

We examine and compare the standard seismic inversion methodology, denoted SM, and full-waveform inversion, denoted FWI. We find many parallels but also interesting differences. Both methods produce a detailed impedance model (or impedance image) as the end product but differ in how this is created. SM first produces a reflectivity image (i.e. a migrated section) that is then converted to impedance, in a step called impedance inversion, by incorporating low-frequency information from an external source, typically well control. In a preparatory step, the reflectivity image is calibrated by comparing it to synthetic seismograms at well locations. We call this well validation and it serves to estimate the seismic wavelet whose removal matches the seismic reflectivity image to the well reflectivity. Alternatively, FWI creates an impedance image as the result of an iteration which gradually adds detail into an initial impedance model. The impedance update at each iteration comes from a type of migration of the data difference, which is the difference between the recorded data and synthetic data predicted by the impedance model as it exists at the iteration's beginning. This migrated data difference is derived from theory as the gradient of the data misfit function, or sum-of-squares of the data difference. Essentially the impedance model is calibrated by comparing synthetic data to recorded data, and we call this data validation.

Both methods require low frequency information but FWI requires this in the data while SM incorporates wells. Both methods require knowledge of the source waveform, but SM achieves this by deconvolution and tying to wells which FWI commonly estimates this in the iteration. SM validates the model at wells and never attempts to predict synthetic data. FWI validates the model through data prediction and comparison to the raw data. SM produces a migrated reflectivity image while FWI uses migration to estimate the gradient of the misfit function. However, we show that this gradient is actually a rather poor migration which lacks gain correction.

FWI is the method of the future but we suggest that a viable step forward is iterative modelling, migration, and inversion or IMMI. Such an approach can incorporate any migration method and can use both well validation and data validation.

INTRODUCTION

Seismic imaging is an essential technology that underlies the exploration for and development of the earth's hydrocarbon resources. Despite this essential role and over 50 years of evolution, the technology is far from optimal and is still rapidly improving. Today's seismic images are vastly superior to those of even 20 years ago and techniques have been developed to extract far more information from them. These images are constructed by passing the raw data through a standard methodology (SM) of data processing (e.g. Yilmaz 1987) that, while very technologically sophisticated, is derived from a complex blend of physical theory and practical experience that has evolved over the past 50-60 years. Now, a new comprehensive inversion process is emerging which is more firmly rooted in modern mathematical physics and inverse theory. Called *full-waveform inversion* or *FWI*, (e.g. Lailly 1983, Tarantola 1984, Pratt 1999, Virieux and Operto 2009) this relatively newer technique promises to deliver more accurate estimates of subsurface properties than SM but that promise is far from realized. This paper examines the underlying principles of both methods with the intent to uncover commonalities and linkages. In particular, we make a detailed examination of imaging conditions as used in prestack migrations in SM and also in the gradient calculation of FWI. We show that the FWI gradient is a particularly poor migration which lacks gain correction. Furthermore we argue that replacing the FWI gradient with a prestack depth migration using a deconvolution imaging condition is similar to estimating and applying the inverse Hessian to the FWI gradient. We also examine two alternate forms of earth model validation: well validation as used in SM and data validation as used in FWI. We argue that both are valuable and both should be used in ways that leverage off their different strengths. The possibility of progression from SM towards FWI through the incorporation of techniques from SM into the FWI paradigm will be explored. We suggest the term *Iterative Modelling Migration and Inversion* or IMMI as a descriptor for such new methods.

DESCRIPTION OF THE METHODS

The Standard Methodology

For the purposes of this discussion, the standard methodology (SM) will be abstracted to a 4 step process of

(1) compensation for attenuation (anelasticity)
(2) estimation and removal of the source waveform (deconvolution)
(3) spatial focusing (prestack depth migration)
(4) impedance inversion.

Of course, there are many more steps in any practical sequence such as statics, surface consistent methods, noise filtering, velocity analysis, etc. The definition of SM here is simplified to emphasize the steps most relevant to the present topic.

Steps 1-3 accomplish the construction of a *reflectivity image* which is commonly a bandlimited estimate of normal incidence reflectivity (if stacked) or offset reflectivity (if unstacked). These steps include correction for geometrical wave spreading, reversal of anelastic attenuation, estimation and inversion of the source signature, and spatial

focusing of the data. A great deal of physical theory is involved in these steps including wave propagation and anelastic loss mechanisms.

Step 4 is almost always accomplished with the assistance of well information. The final impedance model, which is a type of earth model, is constructed from the reflectivity image and validated by comparing the seismic reflectivity to estimates from suitable well logs. I will refer to this form of model validation as *well validation*. Roughly, impedance I and reflectivity R are related through $R = \frac{1}{2}d(ln(I))$ (d means differential) so the estimation of impedance involves integration followed by exponentiation of the reflectivity estimate. In the Fourier domain, this integration is a division by frequency (the Fourier variable) which illustrates the importance of low frequencies in seismic imaging. Since division by a small number (i.e. the low frequency) produces a greatly amplified result, the low-frequencies become essential in step 4, but were almost ignorable in steps 1-3. The very lowest frequency is effectively the integration constant or initial impedance and slightly higher frequencies describe the impedance trend. A reflectivity image lacking the lowest frequencies can only estimate impedance fluctuations away from an unknown trend. Unfortunately, and this is a major limiting factor form both SM and FWI, seismic data is normally deficient in frequencies below 10 Hz so step 4 has usually obtained this low-frequency information by external input, most often from well data. This incorporation of well control is done by computing the reflectivity function directly from suitable well logs and comparing these to the SM reflectivity estimate at the well location(s). Many methods have been developed for this comparison but all result in an estimate of a seismic wavelet whose removal gives the optimal reflectivity match and a low frequency trend to add to the seismic impedance image. This well validation (see Alfaraj et al, 2010, and White and Simm, 2003, for a modern overview) is an essential step in exploration and lends great confidence to subsequent drilling decisions. Other limiting factors for SM and FWI are the ever increasing volume of seismic data (and the consequentially large computational burden) and the essentially unknown, highly complex, structure of the earth between the surface and the exploration target.

Full-waveform Inversion

For full-waveform inversion (FWI), the goal is to estimate an earth model such that synthetic seismic data simulated from this model is an optimal match to the observed seismic data. This alternative process of validating the earth model we will call *data validation* and it contrasts dramatically with the well validation used in SM. The theory of FWI originates with Lailly (1983) and Tarantola (1984) and has been evolved extensively since then (e.g. Pratt, 1999). An overview is provided in Virieux and Operto (2009). The match between synthetic and observed data is usually determined by minimizing an objective function which is the sum of squares of the data difference (the subtraction of modelled data from observed data). A key enabler is the theoretical result that the gradient of the objective function (with respect to the earth model) is estimated by a type of reverse time migration (hereafter RTM) of the data difference (hereafter RTM_{DD}). This leads directly to a steepest descent iteration where an initial earth model m_0 (in the simplest case this is a model of P-wave velocity) is updated by $m = m_0 + \lambda RTM_{DD}$ where λ is a scalar step length that must be estimated, usually by a

line search. Alternative update schemes that are more complex are of the form $m = m_0 + H^{-1}RTM_{DD}$ where H^{-1} is either an exact or approximate *inverse Hessian* operator (e.g. Pratt, 1999; Shin et al., 2001; Virieux and Operto, 2009).

Figure 1 shows an abstraction of FWI, with a cycle of four steps being used to indicate an iteration in a steepest-descent method. The figure is made with assuming the simplest possible case of a constant density acoustic inversion for velocity only. Inputs to the iteration are at (a) and (b) and are the initial velocity model and the recorded data. Step 1 is to create synthetic shot records from the initial velocity model having the same source and receiver locations as the recorded data. In step 2, the data difference (recorded data minus synthetic data) is sent through a prestack migration and stacked. This is regarded as a calculation of "the gradient" because, as mentioned above, theory shows the gradient of the misfit function to be a type of RTM migration. Then, step 3 is the determination of the step-length scalar, λ, which scales the gradient into a velocity perturbation. The conventional approach is to use a line search (a 1-D scan over likely values of λ) to find the optimal value. Margrave et al. (2010) proposed using well control to estimate the best λ. Finally, in step 4, the velocity perturbation is added to the initial model to form the updated velocity model which is input into the next iteration.

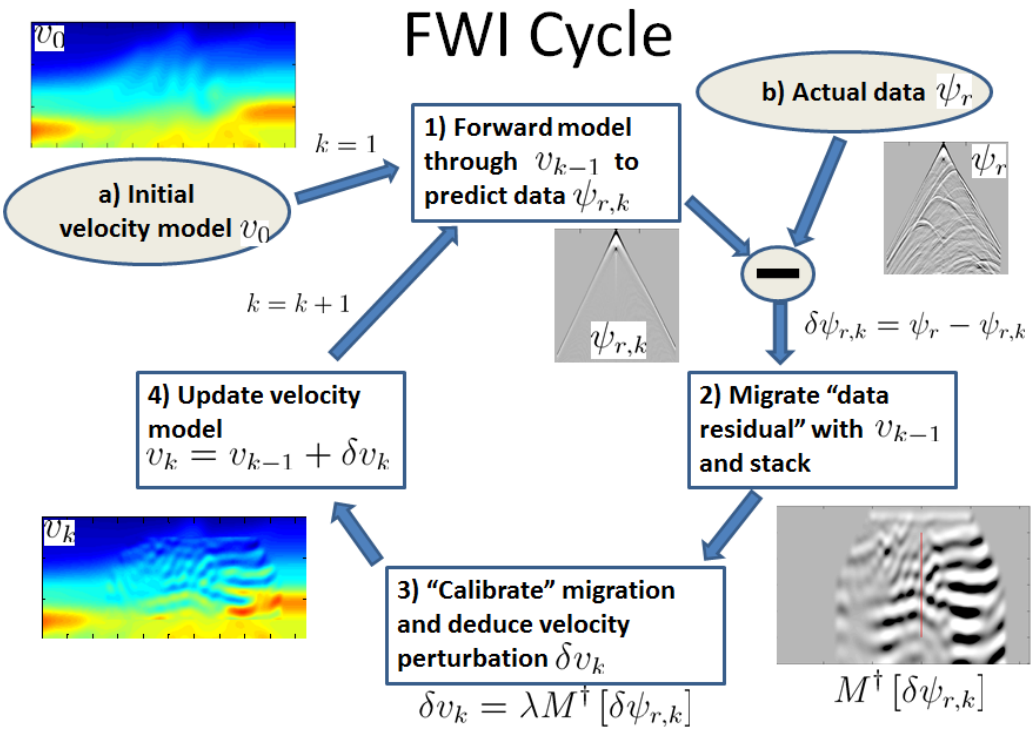

Figure 1: The FWI cycle is depicted for a constant density acoustic inversion. Each iteration of FWI consists of cycling through the four steps shown. Input into each cycle are the current velocity model and the recorded data. The result from each cycle is an updated velocity model (step 4). For the k^{th} iteration, steps 1-3 calculate velocity perturbation, δv_k, as a prestack migration of the data difference, $M^\dagger\left[\delta\psi_{r,k}\right]$.

A major concern in FWI is that, since it is solving a nonlinear inverse problem, there is a danger that the algorithm may become stuck in a local minimum. To possibly avoid this, Pratt (1999) suggests changing the frequency band from low to high as the iteration proceeds. Only the very lowest frequency (or a small band of frequencies) are used on the first iteration which guarantees that the model perturbation will be smooth. The next iteration then uses slightly higher frequencies and so on. Somewhat surprisingly, Pratt finds that inversions limited to 10 Hz and lower can develop models with detail much finer than that expected from migration. Although there is no proof that the procedure will avoid local minima, it is widely accepted as a robust method.

There are several serious issues not illustrated in Figure 1. First, the source waveform must be known to even begin to model the recorded data. However, the source waveform is never known a priori in seismic exploration and must be estimated somehow. In SM, the vast field of seismic deconvolution and wavelet processing is dedicated to this task. Additionally, well control is often very valuable to fine-tune the wavelet estimate. In FWI, this is typically estimated either by capturing an isolated arrival or by a separate inversion. Second, it is almost certain that in a real data case the physics model used in calculating the synthetic data will only approximate the truth. Even a fully elastic model is only an approximation to the real earth which appears to be both viscous and anisotropic in addition to elastic. Moreover the mechanisms behind the viscous damping and anisotropy are only partly understood. Perhaps most commonly, FWI is attempted in the constant-density acoustic case and the consequences of this with real data are not well understood.

FWI needs low-frequency signal for precisely the same reason that SM does. The lower the signal frequencies in the initial model, the less detail is required in the initial model. In principle, with very low frequencies (less than 1 Hz) a constant model is possible. However, while SM usually obtains this information from well control, most FWI approaches make no attempt to incorporate well information. In this paper, as in Margrave et al. (2010) we propose using well control in a variant of FWI.

While FWI offers an attractive validation of the derived earth model (it produces data that match the observed data) it has many difficulties including: (i) the source waveform must be known or estimated (ii) the forward modelling must contain sufficiently complex physics to replicate the observed data (iii) the seismic data must contain low frequencies, (iv) the initial model must be estimated, (v) the computational burden can be immense, and (vi) convergence of the iteration is rarely achieved. These difficulties mean that the adoption of FWI by industry is still a long way off. This research will focus on developing strategies to address these issues.

ON IMAGING CONDITIONS IN MIGRATION

Migration, as used in step 3 of SM, involves the direct estimation of reflectivity which usually is done via an *imaging condition* attributable to Claerbout (1971). Imaging conditions are an essential ingredient to a migration algorithm and specify precisely how an estimate of reflectivity is made. There are two basic types, deconvolution and correlation, and many variants. The deconvolution imaging condition is the more

physical of the two and states that the deconvolution estimate of reflectivity, $\hat{R}_d$, is given by

$$\hat{R}_d(x,y,z,\omega) = \frac{U(x,y,z,\omega)}{D(x,y,z,\omega)}, \tag{1}$$

where (x,y,z) are the Euclidean coordinates of a point in the earth, ω is temporal frequency, the hat ^ over R denotes the frequency domain, and finally U and D are upgoing and downgoing wavefields. This imaging condition is simply a restatement of the definition of reflection coefficient as the ratio of the scattered wave (U) to the incident wave (D). As the estimate of R_d is posed in the frequency domain and hence complex-valued, a migration program will usually be integrated over all recorded frequencies as

$$R_d(x,y,z) = \int \hat{R}_d(x,y,z,\omega)\, d\omega. \tag{2}$$

The recorded data is assumed to be a measurement of the upgoing wavefield $U(x,y,z=z_0)$ where z_0 denotes the surface where the receivers are placed. The estimate of U at any point in the subsurface ($z > z_0$) is done through some form of wavefield extrapolation of the recorded data, commonly either by downward extrapolation in depth or backward extrapolation in time. This is based on an approximate solution to a wave equation, the choice of which amounts to the choice of a governing physics model. A required ingredient of wavefield extrapolation is a prior estimate of the velocity structure beneath the survey, and such estimates are called the *migration velocity model*. The downgoing wavefield D is an estimate of the wavefield from the seismic source (on land typically a dynamite blast or a vibratory vehicle). Usually the source is modelled as a homogeneous Greens function in a small region around the source location and this is then extrapolated to all possible points in the subsurface. A key point, which is only implicit in equation 1 and hence easy to miss, is that this imaging condition assumes that the source waveform in the recorded data is known and that same waveform is used in the estimate of D. In almost all practical settings, the source waveform is not known and must be estimated. In SM, this is the role of deconvolution and, since such estimates are inherently statistical, it must be assumed that the estimates are imperfect. This means that $\hat{R}_d$ as estimated in equation 1, will differ from the "true" value $\hat{R}_T$ by a complex-valued, frequency-dependent, scalar which we denote as

$$\hat{R}_T(\omega) = \hat{\sigma}(\omega)\hat{R}_d(\omega). \tag{3}$$

In a reflectivity image with a vertical time coordinate, as is commonly produced, equation 3 is taken to imply the convolutional relationship $R_T(\tau) = \sigma(\tau) \bullet R_d(\tau)$ where $\sigma(\tau)$ is considered as a *residual* source wavelet that must be estimated and removed. There are other problems with this estimate such as the fact that reflection coefficients in theory are strictly only defined for planar interfaces. We ignore these further complications for the present discussion.

The division involved in equation 1 is computationally problematic because the denominator can be very small at many places in the earth. At such places, the numerator or scattered field will be even smaller and its estimate from recorded data will likely be noise dominated. Hence the division of a noisy estimate by a small number will be wildly inaccurate. A common fix for this is Claerbout's correlation imaging condition and the resulting estimate

$$\hat{R}_c(x,y,z,\omega) = U_g(x,y,z,\omega) D^*(x,y,z,\omega), \qquad (4)$$

where * indicates the complex conjugate and the subscript g denotes that the data (U_g) have been corrected for spherical spreading or *gained*.

A thought experiment to compare imaging conditions

To appreciate the reason for the gain correction, and to obtain a better understanding of the estimates produced by these imaging conditions, we examine a simple thought experiment.

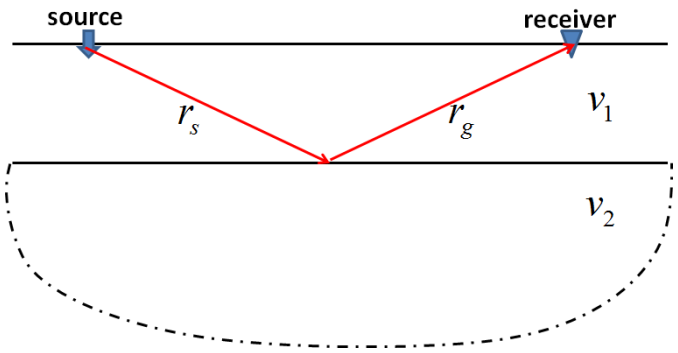

Figure 2: A simple model of a homogeneous layer over a half space. The scalar wave speed in the layer is v_1 and in the half space v_2. The red lines denote the raypath from source to receiver for the specular reflection from the top of the half space.

Consider the propagation of scalar waves in an earth consisting of a horizontal homogeneous layer over a half space (Figure 2). Let the migration velocity model be v_1 everywhere and the migration problem is to estimate the reflection coefficient at the layer bottom. For small angles (assuming offset is much less than depth) the reflection coefficient can be approximated as

$$R_T \approx \frac{v_2 - v_1}{v_2 + v_1} \approx \frac{\Delta v}{2v_1}, \qquad (5)$$

where the second approximation follows if $v_2 - v_1 \equiv \Delta v$ is small. The reflectivity, $\Re$, is then taken to be a function zero everywhere except at the layer bottom where it equals R_T. Assuming the Helmholtz equation (i.e. the scalar wave equation after Fourier transformation over time) applies, the Helmholz Greens function for a homogeneous unbounded medium can be used to model propagation and it is

$$g(r,\omega)=\frac{1}{4\pi r}e^{ikr}\,, \tag{6}$$

where $r=\sqrt{x^2+y^2+z^2}$ is the distance from a unit strength source at the origin and $k=\omega/v_1$ is the wavenumber in the layer. Letting r_s be the distance from the source to an arbitrary point in Figure 2, then migration will estimate the downgoing field as

$$D=W(\omega)g(r_s,\omega)=\frac{W(\omega)}{4\pi r_s}e^{ikr_s}\,, \tag{7}$$

where $W(\omega)$ is the presumed source waveform. Similarly, letting r_g be the distance from the receiver to the same point, then the upgoing field at the receiver is

$$U(z=z_0)=\Re W_T(\omega)g(r_s+r_g,\omega)=\frac{\Re W_T(\omega)}{4\pi(r_s+r_g)}e^{ik(r_s+r_g)}\,, \tag{8}$$

where $W_T(\omega)$ is the actual source waveform. Note that for the specular reflection we have $r_s=r_g$ and $\Re=R_T$ so that

$$U(z=z_0)_{specular}=\frac{R_T W_T(\omega)}{8\pi r_s}e^{i2kr_s}\,. \tag{9}$$

Concentrating on the estimate of the specular reflection, migration will "downward continue" equation 9 to the reflection point at z_r to obtain

$$U(z=z_R)_{specular}=\frac{R_T W_T(\omega)}{4\pi r_s}e^{ikr_s} \tag{10}$$

Comparing 9 and 10 shows that downward continuation is accomplished by

$$U(z=z_0)_{specular}=U(z=z_R)_{specular}\,r_s e^{-ikr_s}\,. \tag{11}$$

We can now insert equations 7 and 10 into the imaging condition 1 to obtain

$$\hat{R}_d(z=z_R,\omega)=\frac{U(z=z_R,\omega)}{D(z=z_R,\omega)}=\frac{R_T W_T(\omega)}{4\pi r_s}e^{ikr_s}\left(\frac{W(\omega)}{4\pi r_s}e^{ikr_s}\right)^{-1}$$

or

$$\hat{R}_d(z=z_R,\omega)=\frac{W_T(\omega)}{W(\omega)}R_T\,. \tag{12}$$

Thus, if $W_T=W$, then we correctly estimate the specular reflection coefficient, otherwise the estimate is off by a complex valued scalar as in equation 3.

Now, consider the correlation imaging condition of equation 4. Applying gain to equation 9, gives

$$U_g\left(z=z_0\right)_{specular}=2r_sU\left(z=z_0\right)_{specular}=\frac{R_TW_T\left(\omega\right)}{4\pi}e^{i2kr_s} \tag{13}$$

and downward continuation to the reflector depth gives

$$U_g\left(z=z_r\right)_{specular}=r_se^{-ikr_s}U_g\left(z=z_0\right)_{specular}=\frac{R_TW_T\left(\omega\right)r_s}{4\pi}e^{ikr_s}. \tag{14}$$

Inserting equations 7 and 14 into 4 gives

$$\hat{R}_c(z=z_r,\omega)=U_g\left(z=z_r,\omega\right)_{specular}D^*\left(z=z_r,\omega\right)=\frac{R_TW_T\left(\omega\right)r_s}{4\pi}e^{ikr_s}\frac{W^*\left(\omega\right)}{4\pi r_s}e^{-ikr_s}$$

or

$$\hat{R}_c(z=z_r,\omega)=\frac{W_T\left(\omega\right)W^*\left(\omega\right)}{\left(4\pi\right)^2}R_T. \tag{15}$$

The estimate in equation 15 is comparable to that in 12 and is achieved without a risky spectral division. Of further interest for the discussion on FWI, is to consider what happens in the correlation imaging condition if the data are not gained. This gives

$$\hat{R}_{c\_ungained}(z=z_r,\omega)=U\left(z=z_r,\omega\right)_{specular}D^*\left(z=z_r,\omega\right)=\frac{R_TW_T\left(\omega\right)}{4\pi r_s}e^{ikr_s}\frac{W^*\left(\omega\right)}{4\pi r_s}e^{-ikr_s}$$

or

$$\hat{R}_{c\_ungained}(z=z_r,\omega)=\frac{1}{r_s^2}\frac{W_T\left(\omega\right)W^*\left(\omega\right)}{\left(4\pi\right)^2}R_T \tag{16}$$

So the estimate is off by a spatially variant factor of r_s^{-2} which is a much more serious error than in equations 12 or 15. As will be shown, this analysis is relevant to FWI.

Finally, we remark that a robust deconvolution imaging condition may be formulated as

$$\hat{R}_d{}'(x,y,z,\omega)=\frac{U\left(x,y,z,\omega\right)D^*\left(x,y,z,\omega\right)}{D\left(x,y,z,\omega\right)D^*\left(x,y,z,\omega\right)+\mu D_{\max}\left(\omega\right)}, \tag{17}$$

where $D_{\max}\left(\omega\right)=\underbrace{\max}_{x,y,z}\left[D\left(x,y,z,\omega\right)D^*\left(x,y,z,\omega\right)\right]$ and $0<\mu\ll 1$ is a small positive number. The quantity $D\left(x,y,z,\omega\right)D^*\left(x,y,z,\omega\right)$ is known as the illumination and, when the illumination is good (i.e. much greater than $\mu D_{\max}$), we have $\hat{R}_d{}'\approx\hat{R}_d$, while when the illumination is weak $\hat{R}_d{}'$ approaches an ungained correlation estimate. Imaging conditions like 17 are quite practical, although the choice of μ may require experimentation.

EXAMINING THE FWI GRADIENT

As mentioned previously, FWI seeks the earth model that produces the best synthetic data. The "best" synthetic is measured by computing and minimizing the sum-of-squares of the data difference, that is

$$\phi_k(\omega) = \sum_{s,r} \left(\hat{\psi}_s(x_r, z=0, \omega) - \hat{\psi}'_{s,k}(x_r, z=0, \omega) \right)^2, \tag{18}$$

where $\psi_s(x_r, z=0, \omega)$ is the s^{th} recorded shot record at frequency ω with receiver locations at x_r, $\psi'_{s,k}(x_r, z=0, \omega)$ is the corresponding synthetic shot record at the k^{th} iteration, and s, r, k are all positive integers. The sum over shots in equation 18 is not strictly necessary since each shot is an independent physical experiment and could be treated separately. Also, minimizing each shot independently would also minimize the sum. However, it is usually regarded as simpler to treat equation 18 as the objective function. The gradient of equation 18 with respect to the model is a hugely dimensional vector, having an independent dimension for each point in the earth model and each parameter. That is, a discrete 2D model, measuring 1000x1000 points and estimating a single parameter at each point is a vector with 10^6 dimensions. It has been shown many times, first by Lailly (1983) and Tarantola (1984), that the gradient is given by

$$\nabla_v \phi_k(x, z, \omega) = \omega^2 \sum_{s,r} \hat{\psi}_{s,k}(x, z, \omega) \delta\hat{\psi}^*_{r(s),k}(x, z, \omega), \tag{19}$$

where $\hat{\psi}_s(x, z, \omega)$ is a modelled source record at subsurface position (x, z) and frequency ω, $\delta\hat{\psi}^*_{r(s),k}(x, z, \omega)$ is the data difference (conjugated), formed at the surface as $\delta\hat{\psi}_{r(s),k}(x, z=0, \omega) = \hat{\psi}_s(x_r, z=0, \omega) - \hat{\psi}'_{s,k}(x_r, z=0, \omega)$, and then back propagated to all possible subsurface locations (x, z), and the * indicates complex conjugation.

While equation 19 is written in the frequency domain, when inverse transformed to time it is recognized as a reverse-time migration (RTM) of time-differentiated wavefields. Comparing equation 19 to the correlation imaging condition of equation 4 shows that the gradient involves a correlation imaging condition (the placement of the * does not matter once the gradient is summed over positive and negative frequencies). However, the data difference is ungained and, from the subsection "Thought experiment to compare imaging conditions" we can anticipate that this will lead to an estimate that decays with depth. Consider the model in that section and the first iteration using a homogeneous initial model. The synthetic data will contain only a direct arrival and so the back propagated data difference will contain only the reflection and is given by equation 10, which we re-write as

$$\delta\hat{\psi}^*_{r(s),1}(x, z, \omega) = \left(\frac{R_T W_T(\omega)}{4\pi r_s} e^{ikr_s} \right)^* = \frac{R_T W_T^*(\omega)}{4\pi r_s} e^{-ikr_s}, \tag{20}$$

and $\hat{\psi}_s(x, z, \omega)$ is given by equation 7 as

$$\hat{\psi}_{s,1}(x,z,\omega)=\frac{W(\omega)}{4\pi r_s}e^{ikr_s}. \tag{21}$$

Then, assuming only 1 source and 1 receiver, the gradient becomes

$$\nabla_v\phi_1(x,z,\omega)=\omega^2\frac{W(\omega)}{4\pi r_s}e^{ikr_s}\frac{R_T W_T^*(\omega)}{4\pi r_s}e^{-ikr_s}=\omega^2 R_T\frac{W(\omega)W_T^*(\omega)}{16\pi^2 r_s^2}, \tag{22}$$

so we see that the gradient does indeed decay as r_s^{-2} in this simples example but we expect similar behavior in more complex models. This is a restatement of the point noted previously that the use of a correlation imaging condition requires that a gain correction be done. The lack of a gain correction here presents serious difficulties for a descent method which seeks to choose a scalar, λ, such that

$$\delta v_1=\lambda\int\nabla_v\phi_1(x,z,\omega)\,d\omega \tag{23}$$

is an "optimal" update to the starting velocity model (here the integral is taken over whatever frequency band is being used in the present step and includes both positive and negative frequencies). Given that the gradient decays as r_s^{-2}, this is clearly a very crude update. Since the velocity model is an earth property, δv_1 should be independent of distance from the source; however, this is clearly not the case. The best that can be hoped for is that λ will represent some sort of average gain correction for the entire section.

Based on the forgoing discussion, it seems appropriate to find a way to incorporate some sort of gain correction into the calculation of the velocity update. The poorly-scaled nature of the gradient was commented upon by Gray (1997) in the context of lease-squares migration and in more detail by Shin et al. (2001) in their insightful analysis of the inverse Hessian. They showed that the diagonal terms of the inverse Hessian apply a gain to the gradient estimation. To understand this better, recall that a steepest-descent method is an iterative approach to an inversion that is generally solved more directly with a Newton or Gauss-Newton approach. The difference in these is that the former calculates the full inverse Hessian operator while the latter calculates an approximation. Mathematically, these methods replace equation 23

$$\delta v_1=\int H^{-1}\nabla_v\phi_1(x,z,\omega)\,d\omega \tag{24}$$

where H^{-1} is either the full inverse Hessian operator or an approximation. Shin et al (2001) show that the main effect of H^{-1} is a gain adjustment. In a discrete setting, H^{-1} is a matrix whose size is the square of the size of the gradient. So, for a 2D 1000x1000 earth model, H^{-1} is a matrix of size 10^6x10^6 and thus has one trillion entries. Clearly some approximation is needed. Using just the diagonal terms of H^{-1}, which can be viewed as an image the same size as the gradient, Shin et al (2001) produced the results in Figure 3. Figure 3a represents the model to be estimated by updating a smooth initial model. Figure 3b is the estimated gradient obtained from an implementation of equation 19 and the loss of amplitude with depth is clearly evident. In Figure 3c is show the diagonal terms of the H^{-1} formed into an image the same size as the gradient. The division of each point of b) by the corresponding point of c) is an approximate

implementation of equation 24 and the result is in d). Clearly, this is a better velocity update than a simple constant-scaled version of b).

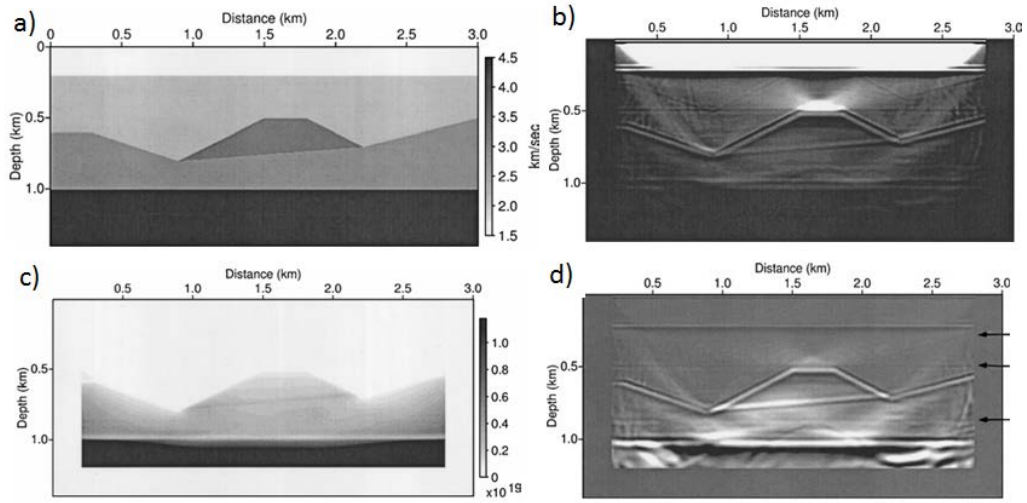

Figure 3: Adapted from Shin et al (2001). a) The "true" velocity model. b) The gradient estimate resulting from an initial velocity model and equation 19. Note the gradual fade-out with depth. c) The image formed from the diagonal entries of H^{-1}. d) The pointwise division of b) by c) representing equation 24.

Last year we suggested (Margrave et al, 2011a) that a conventional prestack migration using a deconvolution imaging condition accomplished a result similar to that of Shin et al. (2001) but with far less computational effort. However, our analysis was purely theoretical and the final result seems to contain an extra factor of r^2, which we have not yet resolved. Here, we present a numerical result using the Marmousi model in the context of the numerical experiment presented in Margrave et al (2010). Figure 4a shows the gradient estimation made using equation 19 for a migration with the Marmousi model. For this experiment only 41 shots were used, placed regularly from 4000-800m along the model. For the result in Figure 4, only frequencies from 0-5 Hz were used and the initial model was a heavily smoothed version of the true model (see Figure 5). Examination of Figure 4a), shows two undesirable features: the slow decay with depth and a dark band at the top. The former has already been discussed, while the latter is an artifact resulting from incomplete cancellation of the direct arrival when forming the data difference (a similar artifact may be present in Figure 3b.) In Figure 4b), the migrated shot records of the data difference were gained before stacking. This is a better estimate of an earth property in that both negative features of the direct gradient have been reduced. The gain correction used in Figure 4b is only approximate as it was accomplished after the migrated by a simple scaling by $r_s^{1/2}$ as is appropriate for 2D. An accurate gain correction requires full modelling of wave propagation and this can be accomplished, with minimal additional cost, by simply implementing a deconvolution imaging condition in the migration instead of a correlation condition. The result of this is shown in Figure 4c). Also implemented in this result is an ordinary stacking mute as is commonly used to suppress artifacts caused in the shallow section by low fold. While the

mute may be overly severe and more information could be allowed into the shallow section, such a tool is clearly beneficial given the right parameters. Using the estimate of Figure 4c, and a further calibration by comparison to well control, Figure 5 shows the resulting velocity model update that results.

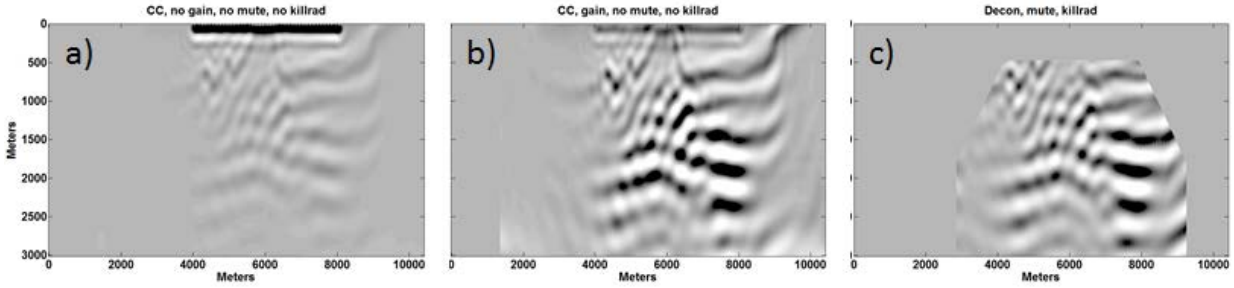

Figure 4: Adapted from the experiment of Margrave et al. (2010). a) The gradient from the first iteration of a FWI using the model of Figure 5b. Note the decay with depth. b) The result of applying gain to each migrated shot record. c) A prestack depth migration using a deconvolution imaging condition.

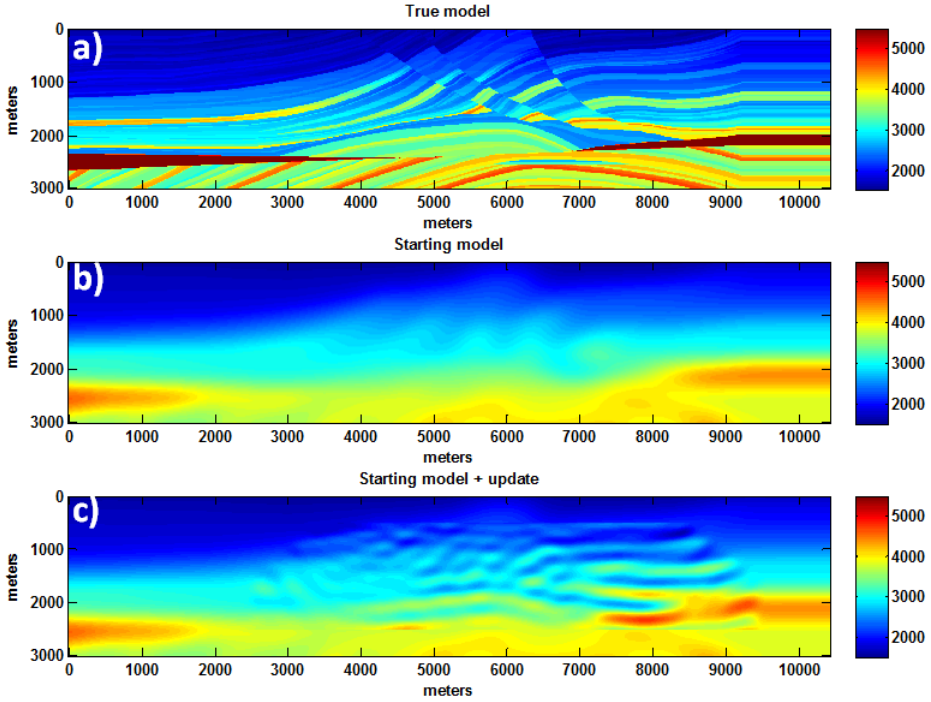

Figure 5: From Margrave et al. (2010). a) The exact Marmousi velocity model. b) The smoothed starting model formed by convolving a 2D Gaussian (half-width 600m) with the exact model. c) The velocity model after updating with the calibrated migration stack of Figure 4c. The update was scaled and phase rotated in a standard comparison with simulated well control.

IMMI: TRANSITIONING FROM SM TO FWI

In the previous section, we have demonstrated that the FWI gradient is a very poor migration. We then showed that simple improvements, using techniques (depth migration

with deconvolution imaging condition and muting) borrowed from SM, improve the gradient and, presumably, lead to faster convergence. We showed that the simple use of a deconvolution imaging condition, which has an almost un-measureable cost increment relative to correlation, accomplishes something very similar to the application of the inverse Hessian, which can be hugely expensive. From this we conclude that there is much to be gained from incorporating aspects of SM into FWI. We suggest that there is a broad middle ground between SM and FWI whose exploration will bring progress from the former to the latter. As a useful name to refer to applications in this middle ground, we suggest IMMI or *Iterative Modelling Migration and Inversion*. Below is an algorithm for IMMI using techniques from both SM and FWI. We present this algorithm for the simplest possible case of constant density, P-wave only estimation of velocity.

1. Prepare the data. If the modelling to be done uses limited physics, for example acoustic modelling for land data, then this dictates at least some of the processing. Ground roll may need to be suppressed and deconvolution may be required. Strongly nonstationary data may benefit from an inverse Q filter or a Gabor deconvolution (e.g. Margrave et al., 2011b). An effort should be made to estimate the wavelet at the end of this step.

2. Build initial background model as a very smooth migration model. The background model should be capable to replicating the data first breaks but should show almost no reflections. This model becomes the current model for the first iteration.

3. Create synthetic seismic data with the current model and the geometry of the real seismic data using the current wavelet estimate. Generally, this will be done with finite difference modelling.

4. Migrate the data difference with a prestack depth migration. Initially migrate only the lowest frequencies. As the iteration proceeds move up the frequency band. We recommend using an f-x migration algorithm like PSPI with a deconvolution imaging condition. Stack the migrated shot records with whatever mute seem appropriate.

5. Convert the migrated stack into a velocity update either through a line search or, if well control is available, by tying the stack to logs. A mixture of both methods may prove useful. If tying to wells, then a wavelet update may also be obtained. The procedure using well control is essentially analogous to standard impedance inversion.

6. Update both velocity model and wavelet and repeat steps 3-6 as often as desired.

Given that SM never iterates while FWI is often iterated hundreds of times, it is reasonable to expect that a dozen iterations will prove very useful. While this suggested IMMI approach is yet untested on real data, its feasibility was shown on the Marmousi synthetic by Margrave et al. 2010.

An Impedance Imaging condition

The deconvolution imaging condition is designed to estimate reflectivity directly. When used in IMMI, the resulting reflectivity image must then be converted to impedance. This conversion could be made easier by using an imaging condition that estimates impedance directly. Such a condition was suggested previously by us in Margrave et al. (2011a). Recall that at interface j, the normal incidence reflection coefficient and impedance are related through

$$R_j = \frac{I_{j+1} - I_j}{I_{j+1} + I_j} \tag{25}$$

where I_j and I_{j+1} are the impedances above and below the interface and R_j is the impedance at the interface. If $I_j \approx I_{j+1}$ then we can write this as

$$R_j = \frac{\Delta I_j}{2I_j} \tag{26}$$

from which we obtain

$$\Delta I_j = 2I_j R_j \,. \tag{27}$$

We propose to treat equation 27 as a prescription to obtain an impedance perturbation from a reflectivity estimate during an IMMI iteration. Specifically, our impedance imaging condition is then

$$\Delta I_k = 2I_{k-1} R_k \tag{28}$$

where k now counts the iteration number in IMMI and I_{k-1} is the impedance model at the start of iteration k, R_k is the reflectivity estimate made iteration k from the deconvolution imaging condition, and ΔI_k is the corresponding impedance perturbation. Thus the updated impedance model will be

$$I_k = I_{k-1} + \Delta I_k \,. \tag{29}$$

If the source wavelet is known then equation 29 can be used directly; however, if the source wavelet is imperfectly known, then equation 28 will need an amplitude and phase correction before using the result in equation 29 (see the discussion around equation 12). It is likely that this correction can be estimated from comparison to well control.

WELL VALIDATION VERSUS DATA VALIDATION

We have discussed two distinct methods of validating an impedance model. In SM, a reflectivity image is compared to well control and then converted to impedance by including the low-frequency contribution from the well. In effect, the resulting impedance section has been validated by comparing to well control, hence we call this well validation. In FWI, an impedance model is validated by demanding that synthetic

seismic data predicted from the impedance model match the real data; so we call this data validation. Both of these approaches also offer opportunities to estimate the seismic wavelet.

While well validation and data validation are both desirable, they are distinctly different constraints, having different features and different computational algorithms and effort. Well validation is typically a relatively minor computational burden in that is requires comparison of the reflectivity image to well control at a small number of spatial locations. In addition, the well logs will usually span only a small fraction of the depth range of the seismic image and almost always there is no logging coverage in the shallow section. Data validation, in contrast, requires solution of a wave-equation initial value problem in order to forward model the synthetic data. While not all measured shot records need be modelled, this is still a much larger computational effort and often requires many iterations of FWI to achieve a good fit.

Usually, seismic migration uses a physical theory that is known to be much simpler than the anisotropic visco-elastic theory that would be needed to fully simulate seismic waves. Most migrations today still use a scalar wave theory, usually both lossless and isotropic, to form an image. In SM, the image is a reflectivity image while in FWI it is a gradient of an objective function, but both are accomplished by migration algorithms with simplified physics. Well validation can proceed with such inadequate physics by simply assuming that the image represents P-P or P-S or other reflectivity and deriving an appropriate wavelet to match the image to the appropriate logs. The issue is far less clear with data validation, but it seems that inadequate physics will have a greater impact because, for example, a scalar wavefield might never achieve a useful match against elastic data. This places an additional burden on data pre-processing as the effects which cannot be modelled should be suppressed.

The two validation methods have different sensitivities to perturbations at different depths. Data validation is far more sensitive to changes in the shallow part of the model than well validation. This sensitivity is directly indicated by the FWI gradient with it's rapid decay with depth. While this decay is not a desirable feature of an earth property, it is a valid statement about the sensitivity of the FWI objective function to changes in the shallow section. On the other hand, well validation is limited by the depth range of the logging. There will always be an unlogged portion in the upper section due to practical considerations. Generally, in a well-tying exercise, this portion is filled with a simple constant gradient function that gives roughly correct traveltimes. However, well tying is very definitive in the logged interval and can effectively determine both the seismic wavelet and the scaling from reflectivity to impedance.

CONCLUSIONS

We have argued that there is a close connection between the standard seismic processing methodology (SM) and the emerging technique of full waveform inversion (FWI). Both methods have the goal of creating an earth model, in the present context an impedance model, from seismic data and both require low-frequency signal content and knowledge of the seismic waveform.

SM, being far older and having evolved from early intuitive ideas to more mature physical theory, is a rich collection of algorithms typically run in a sequence with the goal to create a reflectivity image which is then combined with well information to estimate impedance. The process of comparing the reflectivity images to well control offers a type of model conformation which we call *well validation*. The rich algorithmic set in SM is designed to deal with the many complexities of seismic data such as random and coherent noise, anelasticity, near surface complexities and statics, wavelet estimation and deconvolution, anisotropy, spatial focusing (migration), and more. Interestingly, while detailed impedance sections are commonly made, there is never any attempt to predict synthetic data from them or to use them to re-migrate.

FWI shows great promise as a possible step forward into a new imaging paradigm with potentially much greater resolution. The method is formulated as a nonlinear inversion that strives to minimize and objective function that measures the misfit between real and predicted data. The resulting impedance model is thus consistent with the recorded data and we refer to this as *data validation.* While data validation is a very strong constraint, the method takes much greater computational effort than SM and places a much stronger demand on properly modelling the physics of seismic waves.

We showed that the FWI gradient, which is often said to be a reverse time migration (RTM) is actually a very poor migration as it produces an ungained image. This is indicative of the fact that changes in the near surface have a greater effect on the data misfit than deeper changes, but it is also a likely reason that FWI convergence is slow. A gained image can easily be achieved by replacing the correlation imaging condition inherent in the FWI gradient with a deconvolution imaging condition. We argued that this is a very close parallel to the correction obtained by applying the inverse Hessian operator to the gradient. However, the deconvolution imaging condition is far less computational effort than estimating the inverse Hessian.

Therefore, we suggest that an FWI like process, which we call *iterative modelling migration and inversion* or IMMI, can be created from standard techniques such as finite difference modelling, any standard prestack depth migration of the data difference, calibration of the migrated section using well control to deduce an impedance update, and iteration. This has been demonstrated with synthetic data and is a focus of current research with real data.

ACKNOWLEDGEMENTS

We thank the sponsors of the CREWES project and NSERC for their generous support. We acknowledge useful discussions with Rob Ferguson, John Bancroft, Dave Nichols, and Hassan Khaniani.

REFERENCES

Alfaraj, M. N., M. R. Hong, S. A. AL-Dossary, J. Wang, J. L. Rice, 2010, Wavelet extraction assessment for quantitative seismic interpretation: First Break, 28 , no. 3, 65-71.

Bertram, M., B. and Margrave, G. F., 2010, Recovery of low frequency data from 10Hz geophones: in the 22nd Annual Research Report of the CREWES Project.

Claerbout, J., 1971, Towards a unified theory of reflector mapping: GEOPHYSICS, **36**, 467-481.

Gray, S. H., 1997, True-amplitude seismic migration: A comparison of three approaches: Geophysics, 62, 929–936.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society of Industrial and Applied Mathematics, Expanded Abstracts, 206-220.

Margrave, G. F., Ferguson, R. J., and Hogan, C. M., 2010, Full waveform inversion with wave equation migration and well control: in the 22nd Annual Research Report of the CREWES Project.

Margrave, G. F., K. Innanen, and M. Yedlin, 2011a, Full waveform inversion and the inverse Hessian: in the 23nd Annual Research Report of the CREWES Project.

Margrave, G. F., Lamoureux, M. P., and Henley, D. C., 2011b, Gabor deconvolution: Estimating reflectivity by nonstationary deconvolution of seismic data: Geophysics, 76, 15-30.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: GEOPHYSICS, 64, 888-901.

Shin, C., K. Yoon, K. J. Marfurt, K. Park, D.Yang, H. Y. Lim, S. Chung, and S. Shin, 2001, Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion, Geophysics, 66, 1856-1863.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: GEOPHYSICS, 49, 1259-1256.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: GEOPHYSICS, 74, WCC1-WCC26.

White, R., and R. Simm, 2003, Good practice in well ties: First Break, 21, 75–83.

Yilmaz, 1987, Seismic Data Processing, SEG books.