
Chapter 0.1

IMAGING AND INVERSION WITH ACOUSTIC AND ELASTIC WAVES

John A. Scales
Department of Geophysics
Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401
USA

ELASTIC WAVES
waveform inversion, inverse theory, imaging, migration, seismology

Contents

0.1.1	Seismic Imaging	1
0.1.2	Mathematical Formulation	3
0.1.3	Inversion Versus Imaging	9
0.1.4	Elastic Full-Waveform Inversion	10
0.1.5	Conclusions	14
0.1.6	Acknowledgements	14

Imaging the interior of a region with waveforms recorded on the surface is a key application of scattering theory in geophysics, medical imaging and many other fields. In this chapter I will discuss how the waveform inversion problem can be formulated in terms of statistical inference. Statistical methods are essential in order to be able to take into account the limited, uncertain nature of the observations. The goal will not be to find the “true” model of the sampled region (i.e., its boundaries and/or material properties), since there will always be infinitely many models that fit the data, but rather to put bounds on the feasible values of the parameters that characterize the model. While the idea of using inference methods in waveform inversion is not new, as proposals along these lines were made at least as far back as the mid-1980s, field data applications with realistic prior information and uncertainty estimates are rare because of the difficulties in quantifying this information. In addition, it will also be seen that the classical imaging methods of seismology and ultrasonics (such as migration and diffraction tomography) arise as special cases of more general waveform inversion formulae. After giving a general formulation of the waveform inversion problem, I will show an example of the application of these ideas to elastic waveforms from a reflection seismic experiment. While the application shown is drawn from exploration geophysics, the methods described are applicable in many fields that make use of array-based waveform measurements, such as ultrasonics, non-destructive testing, global seismology, ocean acoustics and ground-penetrating radar.

0.1.1. Seismic Imaging

The imaging of subsurface geology with elastic waves is the most widely used application of scattering theory in geophysics. Seismic imaging methods go by various names, depending on the application, including *migration*, *synthetic aperture focusing*, and *diffraction tomography*; we will see shortly that these methods are special cases of more general waveform inversion procedures. Before proceeding to a precise formulation of the waveform inversion problem, I will begin with a little background on seismic imaging and data processing. Experts on seismic data processing may want to skip ahead to the next section.

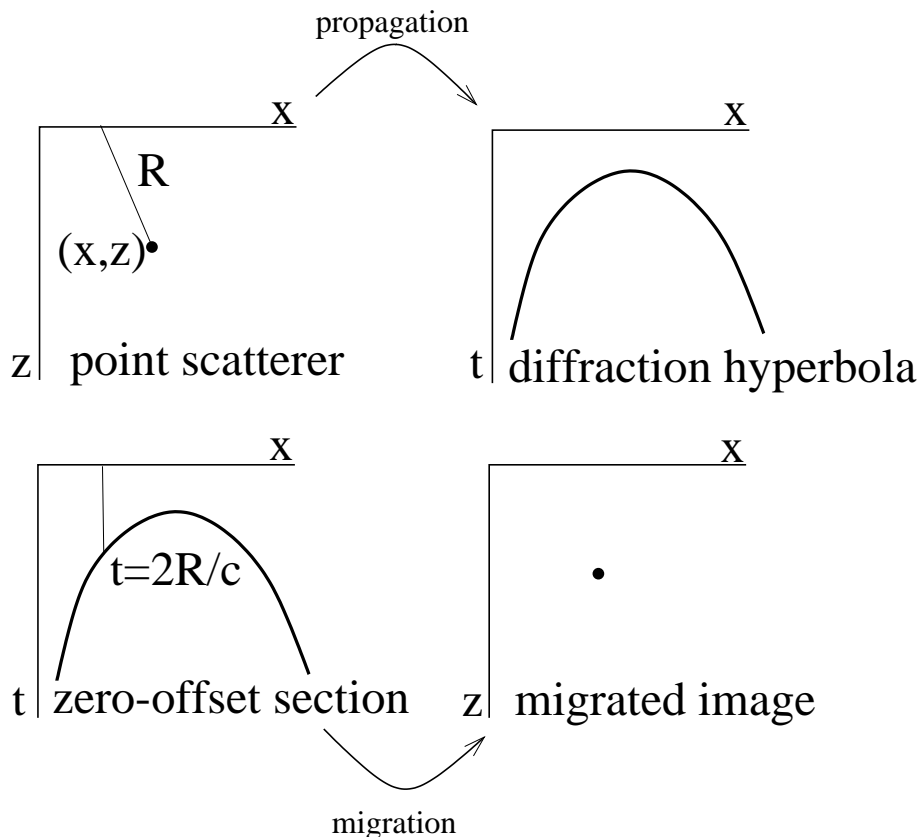
The term *migration* goes back half a century to the earliest uses of seismic reflection profiling in which mechanical means were used to move (or migrate) energy in collections of seismograms recorded as a function of horizontal distance from a controlled source (the *offset*) into its proper location in depth. I will use the term *data domain* to refer to the organization of the data into horizontal distance (offset) versus time. The term *spatial domain* will refer to horizontal distance versus depth. So, one of the results of migration is to transform seismic traces from the data domain to the spatial domain, thus producing an *image* of the medium.

For instance, in a homogeneous medium, a point diffractor in the subsurface produces a hyperbolic pattern in the data domain, whereas the proper spatial image (offset-depth) should be a point. So migration seeks to collapse point-diffraction hyperbolae into (band-limited) points in space (see Figure 1).

By extension, the same argument applies to collections of point scatterers such as reflecting boundaries, cracks, etc. In the modern sense, the term migration is used to refer to any method of producing a map of the subsurface reflection coefficient.

The earliest computerized migration methods were based on sorting and stacking the reflection seismograms. For example, given a collection of seismograms recorded along a line over a horizontal re-

Figure 1 A point diffractor in a constant wavespeed medium produces a hyperbolic pattern in the data domain (horizontal location as a function of time). This is just the Pythagorean theorem. As we will shortly see, one way to invert this process, thereby collapsing the diffraction hyperbola onto a (band-limited) point in the spatial domain (horizontal distance as a function of depth), is to sum the data along isochrons. For each output (image) point (x, z) in the subsurface, sum along the curve defined by $t = 2R/c$, where R is the distance between the image point (x, z) and the receiver location on the surface. The procedure will be formalized later via the Kirchoff integral (or the generalized Radon transform).



flecting boundary, then for all of the traces whose source/receiver midpoint is the same (a *common midpoint* (CMP) gather) the specular reflections will have illuminated the same small region in the subsurface, albeit with different total traveltime. Neglecting any angular dependence of the reflection coefficient, these common midpoint traces can be used to simulate the effects of a hypothetical zero-offset experiment (coincident source/receiver) by simply correcting for kinematical effects. This *normal move-out* correction shifts all the traces to the zero-offset traveltime. Then the traces are averaged to reduce data volume and enhance the signal. The collection of these CMP stacked traces provides a surprisingly effective means of imaging the subsurface if the geology is not too complicated. Further, as we shall see, the fact that they simulate a zero-offset experiment greatly simplifies the migration. (See the book by Yilmaz⁵² for a general discussion of seismic data processing techniques.)

When faced with more complex subsurface geolo-

gy, such simple methods fail. Later it was recognized that a comprehensive picture of the migration procedure could be provided via the wave equation.³¹⁻⁴¹ Suppose that variations in the earth's elastic properties can be divided into two classes. First are smooth variations due to gradual physical processes such as compaction. Smooth variations in wavespeed can be most effectively determined using kinematic information such as traveltimes, for example, by using traveltime tomography.⁷ Second, there are also rapid variations in wavespeed due to faulting, erosion and climactic changes. These rapid spatial variations give rise to reflection and scattering of the propagating waves.²⁵

If the smooth part of the wavespeed is known, then the wave equation can be used to undo the scattering events that gave rise to the reflection in the first place. In effect, if the data are propagated backward in time on the computer to the point at which the scattering event occurred, the result will be a spatial map in which the strength of the image at each point is related to the strength of the scatter-

ing event (the angle-dependent reflection coefficient). [This is completely different than the acoustic time-reversal procedure pioneered by the group of Mathias Fink in Paris^{13 15} in which a physical experiment is performed to transmit a time-reversed version of the recorded wavefield. This has the effect of “undoing” the propagation, focusing the data on the source or on scatterers within the medium, without any knowledge of the material properties of the medium itself; provided time-reversal invariance is satisfied between the data recording and re-transmission. Although we will see shortly that the migration operator is in fact a time-reversal operator, its application requires knowledge of the smooth elastic properties of the medium.]

The idea is illustrated in Figure 2. It shows the reflected waveforms associated with a synthetic zero-offset experiment performed over a synclinal reflecting boundary. (Or these could be the CMP stacked records simulating a zero-offset experiment.) Although the zero-offset data show the shape of the reflecting boundary fairly clearly, the bow-tie shaped triplication in the traveltime curve makes it difficult to interpret the data in the time domain. The migration procedure is used to produce an image of the subsurface in which the locations of the velocity discontinuities are in their proper spatial location. The resulting migration is shown on the right side of Figure 2. For zero-offset data this is easy to do. Since the down-going wave path is the same as the up-going path, if a reflection event occurs at a given two-way time T in the seismogram, then it was produced by a reflection which occurred in the subsurface at time $T/2$. It’s almost as if, using half the wavespeed, that the data were produced by an ensemble of sources located on the reflecting boundary fired off in unison at $T = 0$. (This is called the *exploding reflector model*.¹⁰) So all one needs to do is propagate the zero-offset data backwards in time, using a wavespeed equal to half the true wavespeed, to $T = 0$. The result will be a map of the subsurface reflectivity. Examples of the details of such calculations can be found in textbooks (e.g., Claerbout¹⁰ and Scales³⁹). Strictly speaking, the validity of the CMP stacking procedure assumes flat reflectors and constant velocity down to each reflector; nevertheless it has proven useful in a wide variety of cases where the assumptions are grossly violated.

Our interest here in zero-offset migration is largely pedagogical. Full waveform inversion provides a general framework that obviates much of the discussion. However, there are many modern generalizations of this simple scheme that go under the general heading of Transformation to Zero-Offset. For more details on the data processing issues related to these tech-

niques see the book edited by Hale.²³

A more interesting example shows the effects of aperture and after-migration stacking of images. In Figure 3 can be seen (on the left) synthetic shot (or common-source) records taken over a reflecting boundary composed of three flat segments. The true structure is very similar to the lower right panel. Note that in the time-sections (on the left) three reflections are visible, whereas there is only one reflector (with three facets). The right column of figures shows the corresponding migrations; each shot record is migrated and the resulting images are summed. This is an example of a pre-stack migration, where the imaging (in this case stacking) is performed after the migration. In the shot (data) domain, the individual images are quite different. However, since they all image the same piece of the subsurface, once the migration has been performed the images can be added together. The top row is for a single source, the middle is for three sources and the bottom is for six.

This example also provides a clue to a more general method of velocity estimation than CMP stacking. If we migrate the shots (or perform any sort of pre-stack migration) and then sort the data into common-receiver gathers, then the result will be a map of the subsurface below a particular receiver for many shots. Since there is only one earth, all the shots should produce the same image of a particular point in the subsurface; any variations must be due to errors in the velocity model and to artifacts due to the finite aperture of the data used for imaging. So we can do velocity analysis by varying the velocity in order to make events in the migrated common

Figure 2 A simple, synthetic example of array-based imaging of the type commonly used in exploration seismology. The model consists of a synclinal reflecting boundary in the subsurface. The synthetic zero-offset (coincident source-receiver) data on the left represent waves reflected off the synclinal structure and recorded along the surface of the earth. To achieve a proper image of the syncline one must process the array of data in order to restore the reflected data (time domain) to their proper spatial location (space domain).

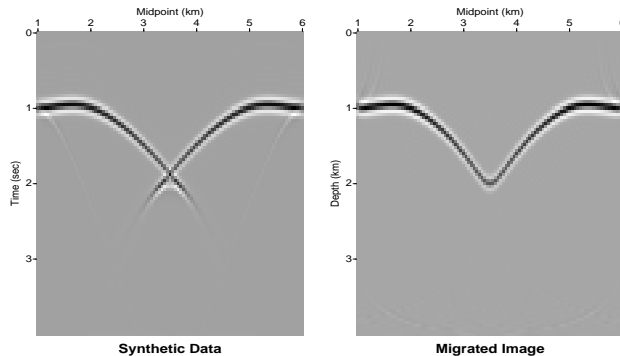
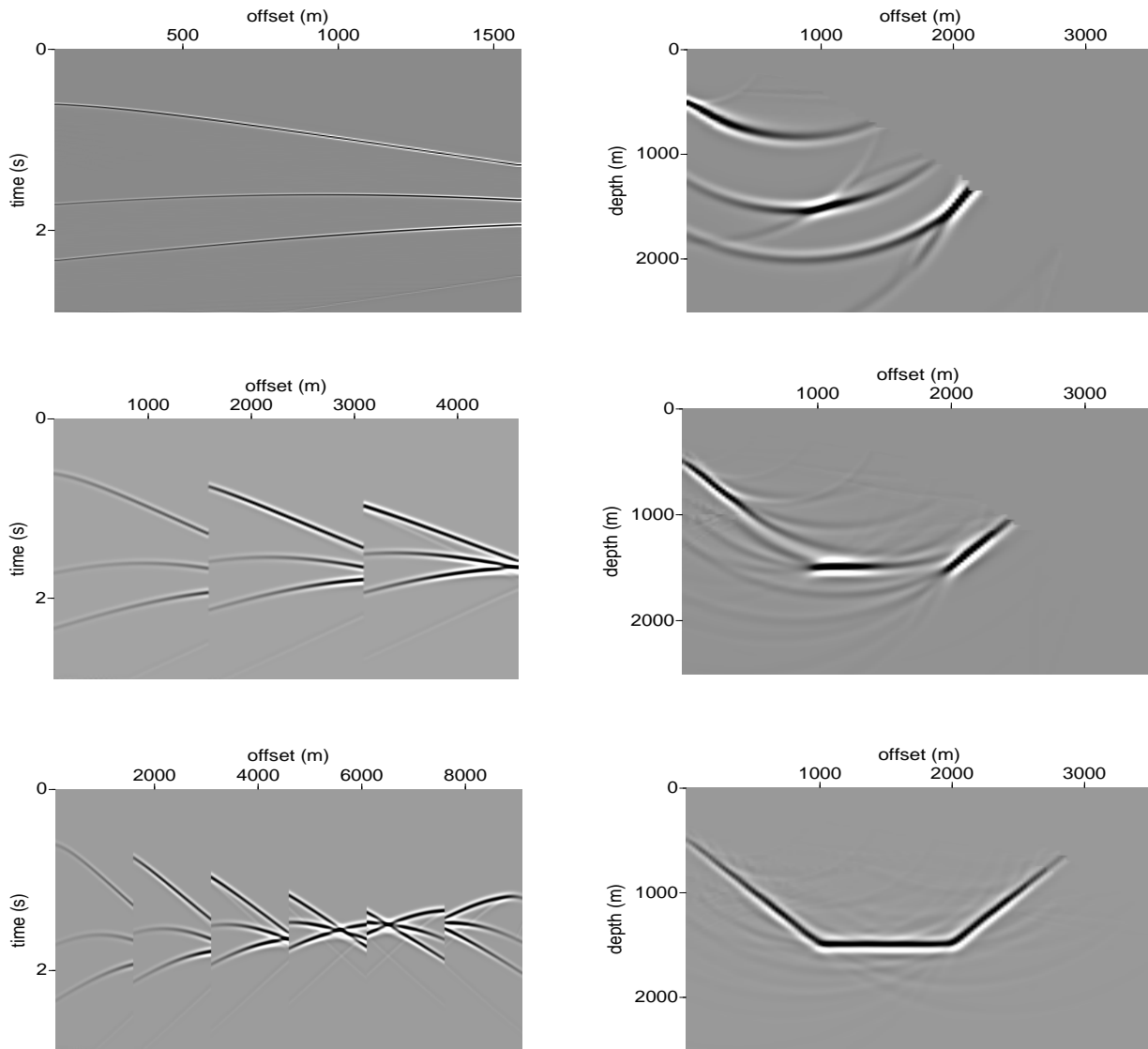


Figure 3 A typical pre-stack migration procedure involves migration of individual shot records or plane waves, which are then stacked to produce the final migrated image of the subsurface. In this simple synthetic example, the geometry of the reflecting boundary, which consists of three flat segments, becomes apparent as the subsurface illumination increases from 1 shot (top) to 3 (middle) to 6 (bottom). Shown on the left are the synthetic data (shot records) on the right, the migrated images. With six shots, a nearly perfect reconstruction of the reflector is achieved.



receiver gathers horizontal. This idea was proposed by Al-Yahya² and forms the basis of most modern velocity analysis schemes.

0.1.2. Mathematical Formulation

The imaging and inversion problems to be considered here begin with a canonical statement of the elastic *forward problem*. Let \mathbf{m} denote a generic element of the space of models \mathcal{M} . A model is a mathematical parameterization of those physical properties that are required to predict the measurements. In gener-

al the model space is an infinite-dimensional space of functions (e.g., mass density as a function of space). The choice of parameters is, of course, not unique. For elastic waves one could use stiffnesses or compliances, velocities or impedances, etc. A good analysis of the issues for elastic waveform inversion is given by Tarantola,⁴⁷ who analyzes this problem by considering the diffraction patterns for isolated perturbations of the different parameters. He shows for P-wave data that at short offsets the diffraction patterns for perturbations in the Lamé parameters are very similar, while for the perturbations in P and

S-wave impedances they are readily distinguished. Therefore, in the field data example described below, impedances will be used.

In practice \mathcal{M} is often replaced by a finite-dimensional vector space. A subtle issue can arise since discretization may result in artificial reduction in the *a posteriori* uncertainty of the parameters.⁴³ For a particular element of model space, the connection between the model and the data is:

$$\mathbf{u}_{\text{obs}} = g(\mathbf{m}) + \mathbf{e} + \mathbf{f}, \quad (1)$$

where \mathbf{u}_{obs} are waveform measurements at discrete receiver locations, g is the data prediction operator (described in detail in the Chapter on Elastic Waves), \mathbf{e} is a vector of random errors, and \mathbf{f} describes any systematic errors (such as unmodeled physics). Evaluating g typically involves solving an initial-boundary value problem of a partial differential equation such as the elastic wave equation. Since any real measurement can consist of only a finite number of data, the space of possible outcomes of the experiment (the data space \mathcal{D}) is always finite-dimensional. The data themselves will be band-limited measurements at discrete receiver locations of one or more components of the elastic displacement, velocity or acceleration (depending on the type of sensor), the electromagnetic field, etc—or they could be scalar measurements of pressure (common in marine acquisition). So, \mathbf{u}_{obs} does not refer to a vector in three-dimensional space, but rather to an element of an abstract n -dimensional space of measurements comprising the totality of samples, components, receiver locations, etc.

The meaning of Equation 1 is as follows: if g is known exactly (this means that it captures all the significant physics of the experiment and can be computed exactly), then in the absence of noise, the true model (i.e., the earth) will predict the data exactly. In practice, even if one knew the true model it would not predict the data exactly because of random and systematic errors. So one must formulate a criterion by which to judge how well a model fits the data.

To keep the theory simple let us restrict attention to a weighted Euclidean mis-fit function of the form

$$\begin{aligned} \chi^2(\mathbf{m}) &\equiv \|g(\mathbf{m}) - \mathbf{u}_{\text{obs}}\|_{C_D^{-1}}^2 \\ &\equiv (g(\mathbf{m}) - \mathbf{u}_{\text{obs}}, C_D^{-1}(g - \mathbf{u}_{\text{obs}})), \end{aligned} \quad (2)$$

where C_D is the data covariance matrix. The symbol (\cdot, \cdot) denotes the inner product in the appropriate space. For elements of data space this always means a summation over the components of the measured time series, as well as summations over all the sources and receivers. The inner product for the models will involve integration in the model space until the prob-

lem is explicitly discretized. This notation hides a lot of messy details but it is easy to get confused about which space the various operations are defined in and whether they are inner products or convolutions.

The data mis-fit, Equation 2, is an appropriate choice if the data uncertainties are normally distributed; it may or may not be appropriate otherwise. The size of the data uncertainties is characterized by C_D . It is up to us to decide how closely the predicted data must match the observations; i.e., how small χ^2 in Equation 2 is to be made. E.g., one would like to find all the models whose χ^2 is less than ϵ . [It may not be necessary to explicitly specify the data error distribution to accomplish this. It may suffice to estimate a confidence set for the errors.⁴²]

Unfortunately because the forward operator g has a nontrivial kernel (null-space), the set of data-fitting models is usually infinite. This is the reason why some data-independent information (called prior information) must be incorporated into the problem. Otherwise it is not generally possible to achieve finite uncertainty on the values of the model parameters. (See Scales and Tenorio⁴⁰ for a discussion of the role of prior information in inverse problems.) For this discussion, I will proceed with a simplified Gaussian/Bayesian treatment and assume the existence of Gaussian *a priori* probability distribution on the space of models. (The term *a priori* simply means independent of the data.) However, the reader should keep in mind that the nature of the Bayesian prior is not obvious in most cases.^{40 37} Often the prior information at our disposal is deterministic (constraints, for example); converting such deterministic information into a probability, even an apparently conservative one, usually results in injecting information into the problem not implied by the constraint.⁴⁰ Further, there are more general formulations of the problem than the Gaussian one, and non-Bayesian methods, too. The literature on the subject is vast. For geophysical applications of inverse theory, see Tarantola⁴⁸ for the Bayesian approach and Parker³⁵ for a frequentist approach. Lehmann and Casella³⁰ give a comprehensive treatment of the statistical ideas that are the foundation for inverse theory.

Using this simplified Gauss/Bayes approach, one can incorporate the prior information alongside the data mis-fit in a combined cost function:

$$S(\mathbf{m}) \equiv \frac{1}{2} \left(\|g(\mathbf{m}) - \mathbf{u}_{\text{obs}}\|_{C_D^{-1}}^2 + \|\mathbf{m} - \mathbf{m}_{\text{prior}}\|_{C_M^{-1}}^2 \right), \quad (3)$$

where $\mathbf{m}_{\text{prior}}$ is the mean of the *a priori* probability distribution and C_M is the covariance matrix of this distribution. In the statistical terminology, $e^{-S(\mathbf{m})}$ is (proportional to) the Bayesian posterior probability distribution that measures the trade-off between

data fit and *a priori* model reasonableness. Models \mathbf{m} that achieve a relatively large value of $e^{-S(\mathbf{m})}$ do a good job of fitting the data and are not too far away from the *a priori* model—distances being measured relative to the covariance matrices. Note that C_M fulfills a fundamentally different role than regularization; while the second term in Equation 3 looks like a regularizing functional, it is not. C_M is supposed to be known *a priori*, so logically it cannot matter whether it regularizes the problem or not.¹⁹

Local Optimization of S

Minimizing Equation 3 is a nonlinear least-squares problem. Notice that the cost function will be quadratic in \mathbf{m} if and only if g is a linear function of \mathbf{m} . In general this is not the case. In particular, for elastic or acoustic waveform inversion, the oscillatory nature of the data means that for models far away from \mathbf{m}_{true} S may be highly non-convex (i.e., have many local extrema). However, it will be seen later that the non-convexity of S comes from the smooth part of elastic wavespeed and not the discontinuities, which are the usual target of imaging methods. In any case the minimization of Equation 3 is amenable to a variety of nonlinear optimization procedures. Most such procedures are based on some sort of iterative linearization of g . At each step, S is replaced by a quadratic function and optimized via a conventional least-squares method. There are different approaches to solving the least-squares problems. On the one hand, a pseudo-inverse solution could be computed at each step, either using direct algorithms such as the singular value decomposition, or iterative algorithms such as conjugate gradient.³⁶³⁸ On the other hand, there may be some advantages to focusing on gradient descent methods since there is a fundamental connection between the gradient of the cost function and classical seismic migration algorithms.⁴⁶⁴⁵

A straightforward calculation (setting the gradient of S equal to zero) shows that in order to minimize S , the model updates must satisfy

$$\mathbf{m} - \mathbf{m}_{\text{prior}} = C_M G^* C_D^{-1} (\mathbf{u}_{\text{obs}} - g(\mathbf{m})), \quad (4)$$

where G^* is the adjoint of the Frechet derivative of g about the prior model. G is defined by

$$g(\mathbf{m} + \delta\mathbf{m}) = g(\mathbf{m}) + G\delta\mathbf{m} + o(\|\delta\mathbf{m}\|^2),$$

while G^* is defined by

$$(G^* \mathbf{d}, \mathbf{m}) = (\mathbf{d}, G\mathbf{m}).$$

Equation 4 shows that the model updates are in the column space of the prior covariance matrix (be-

cause they are linear combinations of the columns). It also suggests the following iteration scheme:

$$\mathbf{m}_{k+1} = C_M G_k^* C_D^{-1} (\mathbf{u}_{\text{obs}} - g(\mathbf{m}_k)) + \mathbf{m}_{\text{prior}}. \quad (5)$$

or, equivalently

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \left\{ C_M G_k^* C_D^{-1} [\mathbf{u}_{\text{obs}} - g(\mathbf{m}_k)] - (\mathbf{m}_k - \mathbf{m}_0) \right\}, \quad (6)$$

where \mathbf{m}_0 is taken to be $\mathbf{m}_{\text{prior}}$. Tarantola⁴⁷ makes this slightly more general by introducing numerical weights:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + W \left\{ C_M G_k^* C_D^{-1} [\mathbf{u}_{\text{obs}} - g(\mathbf{m}_k)] - (\mathbf{m}_k - \mathbf{m}_0) \right\}. \quad (7)$$

W is a purely numerical (preconditioning) weighting operator that has no inverse-theoretic significance; it could be used to improve convergence.⁵⁰

Interpretation

At this point, where we are seeking a conceptual understanding of the waveform inversion procedure, it is worth considering a special case in which the whole procedure is linear (and hence Equation 7 converges after one iteration). This is the so-called convolutional model, in which the data (without multiply-scattered or mode-converted waves) are generated by convolving an operator that depends on the low spatial frequency part of the elastic wavespeed (c_s , “s” for smooth), with the high spatial frequency part c_r (“r” for rough or reflectivity):

$$\mathbf{u}_{\text{obs}} = P(c_s) \cdot c_r.$$

Clearly in this approximation the data depend linearly on the unknown reflectivity. Thus S is a quadratic function of c_r and one iteration of Equation 7 suffices to produce c_r : *if c_s is known*. Based on this one should expect that if the smooth part of the wavespeed is known reasonably well, then a single iteration of gradient optimization would result in an estimate of c_r , which is the goal of migration. So we can see immediately that there is a fundamental connection between migration and linearized waveform inversion.

The dependence of S on the smooth (low-spatial frequency) part of the wavespeed model is more subtle. Changing the background model changes the traveltimes of events in the seismogram. Since the data are highly oscillatory and since S measures the difference between the observed and predicted data, changes in the traveltimes cause the observed and predicted data to go in and out of phase with respect to one another. This creates local extrema in S and means that small changes in c_s can result in large changes in S . Figure 4 shows a simple example. On

the left is a single-source record for a 6-layered acoustic model with a constant wavespeed of 1.5 km/s (using ray theory). The middle shows the source record for a wavespeed of 1.55 km/s. And on the right is the difference between these two; this difference is of the same order of magnitude as the individual sections. So, for a perturbation of less than 5% there would be a substantial change in S .

Let us reduce Equation 7 to its simplest form by setting C_M and C_D equal to identity matrices or operators. (This is wrong of course; not only are the parameters and errors likely to be correlated, but their correlation is likely to depend on space/time. But such a simplifying assumption helps one to interpret the formulae.) We'll also ignore the weighting matrix W .

In this case, performing a single iteration of local optimization on S results in an updated model of the form

$$\mathbf{m}_1 = \mathbf{m}_0 + G_0^* [\mathbf{u}_{\text{obs}} - g(\mathbf{m}_0)]. \quad (8)$$

This equation says that to update the model \mathbf{m}_0 , begin by solving the forward problem (evaluate $g(\mathbf{m}_0)$). This amounts to propagation forward in time of the input wavelet at each source location. The resulting fields are needed at each receiver location. Then compute the data residuals by subtracting the predicted data from the observed data. That will leave the part of the scattered wavefields not predicted by the current model. Next apply the operator G_0^* to the residuals.

The effect of the adjoint operator is to propagate the residual scattered field downward into the subsurface, or equivalently backwards in time. This is clear from Equation 8 since the second term on the right is a model update: G^* acts on data residuals defined at the recording locations and produces a model update everywhere in the subsurface. The incident and scattered fields will be correlated in the subsurface at the location of scatterers and if \mathbf{m}_0 has the correct smooth background, $g(\mathbf{m}_0)$ is just the incident field.

This idea of comparing the forward-propagated source field and the backward-propagated scattered field can be understood with a simple example of a back-propagation/migration procedure.²¹ A vertically incident scalar plane wave reflects off a horizontal boundary at depth $z = z_m$ in a laterally invariant medium. The scattered field at the surface ($z = 0$) (in the frequency domain) is just

$$\psi_S(z = 0) = e^{i\Delta} R(z_m)\psi_I(z = z_m)$$

where ψ_I is the incident field just above the reflector, R is the reflection coefficient, Δ is a phase shift that accounts for the propagation delay down to the reflecting point z_m . We can get rid of the delay on

the right side by back-propagating the scattered field into the subsurface (multiplication by $e^{-i\Delta}$):

$$\psi_S(z_m) = R(z_m)\psi_I(z_m).$$

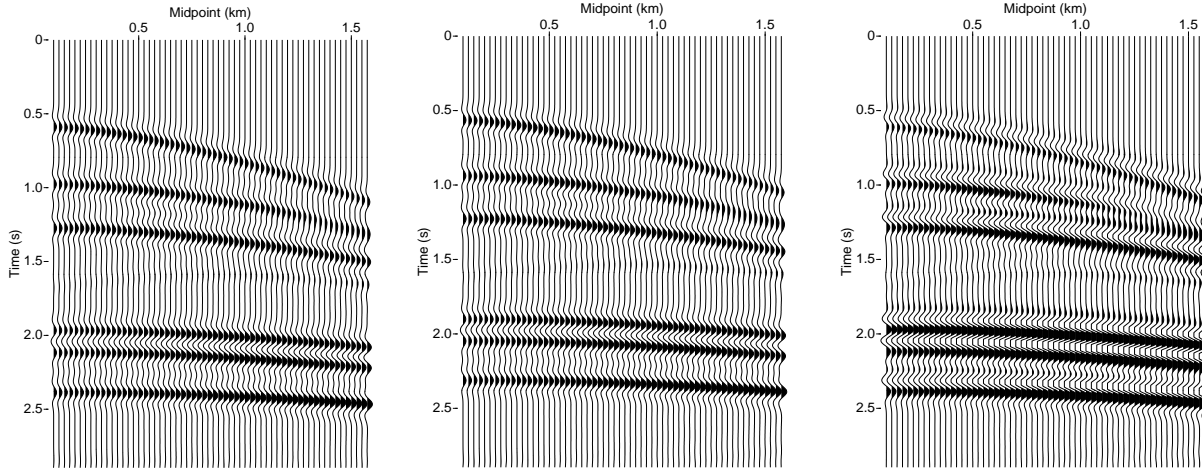
This gives the reflection coefficient at z_m as the ratio of the forward and back-propagated fields. Of course, this idea can be generalized to allow for non-vertical incidence, heterogeneous background, etc. Nevertheless, all these generalizations are special cases of the linearized waveform inversion.

Now, the local optimization formula in Equation 7 gives an inverse-theoretic update to the entire parameter vector \mathbf{m} . If \mathbf{m} consists of the elastic impedance, and if the smooth part of that impedance is known, then the least-squares update for the model perturbation can be thought of as a kind of sophisticated migration procedure, since it produces a map of the jumps in impedance as a function of depth. But instead of dividing the incident and scattered fields at each point in the subsurface, the adjoint operator is used. The detailed interpretation of G^* requires specifying the type of forward modeling and the domain of the migration (common-source, common-offset, zero-offset, etc.). For instance, using a finite-difference method to approximate g yields a linearized inversion algorithm essentially equivalent to finite-difference *reverse-time migration*. Reverse-time migration^{51 24 32} involves reversing the order of the recorded time samples and using them as time-dependent boundary conditions for a finite-difference solution of the elastic wave equation—with time running backwards. On the other hand, using a WKB approximation results in an algorithm essentially equivalent to *Kirchhoff migration*.^{46 26 29} Traditionally, Kirchhoff migration formulae are derived by converting the wave equation into a Fredholm integral equation using a Green's function (or tensor in the vector case).³⁹ Begin with the representation theorem for elastic waves (cf. the Chapter on Elastic Waves or the textbook by Aki and Richards¹).

$$u_i(\mathbf{r}) = \int_{\partial\Omega} G_{in}(\mathbf{r}, \mathbf{r}') n_j c_{n j k l} \partial'_k u_l(\mathbf{r}') - u_n(\mathbf{r}') n_j c_{n j k l} \partial'_k G_{il}(\mathbf{r}, \mathbf{r}') dS', \quad (9)$$

where $\partial\Omega$ is the surface of a closed volume within which there are no sources, c is the rank-4 elastic tensor, and G is a Green's tensor. By definition G is any solution to the wave equation with delta function right-hand-side. If one takes the limit of the observation point as it approaches the boundary, this equation becomes a Fredholm integral equation for the field (or *mutatis mutandis* its normal derivative) on the boundary, given either Neuman or Dirichlet

Figure 4 On the left is a single shot record for a model consisting of six horizontal layers with a constant wavespeed of 1.5 km/s. The middle shows the shot record for the same model but with a wavespeed of 1.55 km/s. On the right, plotted to the same scale, is the difference between these two. Thus, small perturbations in the background velocity model can result in large perturbations in the waveform mis-fit.



boundary conditions.²⁷ At this point the Green's tensor is completely arbitrary, i.e., it need not satisfy any particular boundary conditions.

Instead of attempting to solve this integral equation, it is conventional to reduce it to an integral relation relating the boundary values on the surface to the field in the interior. This is done in two steps. First, the integral over $\partial\Omega$ is reduced to an integral over only the recording surface ($z = 0$ in this case). This requires showing that the integral over the subsurface part of $\partial\Omega$ is negligible. For migration purposes, it is not sufficient to simply take the lower surface to infinity and exploit the radiation condition since for migration one uses an anti-causal Green's tensor which doesn't vanish at infinity. (See Docherty¹⁴ for a stationary-phase argument.)

The second step is to ensure that the Green's tensor satisfies appropriate boundary conditions. For migration, the land-based data are the fields themselves recorded at $z = 0$. (Assuming the recording surface is horizontal; surface topography is another complication.) Thus if the Green's tensor is zero on the surface $z = 0$, the first term on the right-hand side above is zero. These arguments apply to any medium, provided the Green's tensor is known. In the special case of a flat recording surface and a homogeneous medium, an appropriate Green's tensor can be constructed by the method of images.

Assuming these two steps are performed, the representation theorem reduces to the following integral (for Dirichlet conditions):

$$u_i(\mathbf{r}) = - \int_{\partial\Omega} u_n(\mathbf{r}') n_j c_{njkl} \partial'_k G_{il}(\mathbf{r}, \mathbf{r}') dS'.$$

If G is anti-causal and u is the field recorded at the

surface, then this integral gives the migrated image at every point in the subsurface. Although the above argument was given for a homogeneous medium, it still works for smooth models in the WKB approximation provided the traveltimes are properly computed in the smooth background model. Evaluating this integral requires knowing the traveltimes from all the image points in the subsurface to the sources and receivers; in practice these traveltimes are computed by ray tracing or finite-difference solution of the eikonal equation.^{49 22}

Other commonly used forms of migration arise from other approximations to the wave equation. For layered or quasi-layered media, propagator matrices are used (*phase-shift migration*¹⁷). Other examples can be found in Lailly²⁸ and Tarantola^{45 47} who discovered the connection between the local optimization of a least-squares waveform mis-fit function and traditional migration methods.

Whatever method is used to solve the wave equation, the theory of Fourier integral operators can be used to show that *the output of the adjoint, however computed, has high-frequency strength concentrated in the same places as the parameter perturbations.*⁴⁴

Since all uncertainty information has been discarded in Equation 8 the migration model \mathbf{m}_1 cannot be regarded as being the solution of any inverse problem, but it does have some qualitative connection to the discontinuous part of the subsurface elastic parameters. Updated models derived from Equation 8 are referred to as *least-squares migrations*. So, least-squares migration is a poorly scaled (because the covariances are neglected), approximate (because only one iteration is performed) waveform inversion.

Putting the covariance matrices back in, one iteration of Equation 7 gives:

$$\mathbf{m}_1 = \mathbf{m}_0 + W \left\{ C_M G_k^* C_D^{-1} [\mathbf{u}_{\text{obs}} - g(\mathbf{m}_0)] \right\}. \quad (10)$$

So, Equation 10, which is one step of an iterative Gauss/Bayes waveform inversion procedure, can be thought of as the inverse-theoretic generalization of migration. It is always possible of course to discard uncertainty information and simply use Equation 7 as an optimization rather than an inversion procedure. See Chavent and Plessix⁹ for a discussion of an approach to design of optimal least-squares migration procedures.

However, the inversion procedure in Equation 7 is much more general than least-squares migration since:

- g need not be linear; in particular Equation 7 allows for mode-conversion and multiple scattering.
- Uncertainties in the data and *a priori* information on the parameters is explicitly incorporated and the entire procedure results in a combined *a posteriori* estimate of parameter uncertainty. In fact, it is the posterior parameter uncertainty that is the real solution of the inverse problem. The least-squares model may or may not be of particular significance.
- The *a priori* model can be arbitrarily complicated, and it is updated incrementally during the iterative nonlinear inversion procedure.
- The data need not be recorded at regular intervals. The inverse-theoretic formalism allows for redundant and inadequate data.

0.1.3. Inversion Versus Imaging

In the seismic literature the terms *migration*, *imaging* and *inversion* are given various (often contradictory) meanings. The important thing to keep in mind, however, is that imaging and migration are essentially methods of exploratory data analysis. Even though the output of such a calculation is a map of some subsurface property, such as reflectivity or impedance, these maps cannot be quantitatively linked to the Earth’s material properties without performing an uncertainty analysis. Imaging methods are designed to produce pictures that can then be interpreted or used for other means, such as suggesting hypotheses. On the other hand many of the published uses of the term *inversion* suffer the same

drawback because they do not explicitly treat the problem of data uncertainty or the necessity of using *a priori* information.

In order to try to reduce this confusion somewhat, I suggest the following taxonomy:

Imaging Imaging is any method that uses data to produce a model of the subsurface, but where no attempt is made to verify that the model so produced actually fits the data in a rigorous sense or is reasonable *a priori*. All seismic migration fits into this category, including pre-stack “true amplitude” migration.

Construction Finding a model, any model, that fits the data and is *a priori* reasonable is called construction. For example, the output of an imaging algorithm might be shown to fit the data but this requires a careful analysis of the uncertainties of the data. Construction is useful for illustrating the features that the earth *might* have, or for falsifying a hypothesis.

Inference Finally, at the extreme of ambitiousness, one may characterize the entire set of models that fit the data and are *a priori* reasonable. This is the inference problem.

One can think of the inference problem as constructing a box in the space of possible models within which one is certain (to a specified degree) that the true model lies. Before the data have been analyzed, this box might be quite large; without *a priori* information it is infinite. Hopefully, analyzing the data reduces the size of the box. The box might have rigid (deterministic) boundaries, as in frequentist inference, or it might have fuzzy (probabilistic) boundaries, as in Bayesian inference.⁴⁰

Direct inversion methods Finally, there is another class of imaging algorithms sometimes referred to as *direct inversion methods*. These methods seek a direct or analytic inversion of the equation:

$$\mathbf{u}_{\text{obs}} = g(\mathbf{m}). \quad (11)$$

As has already been pointed out this is impossible since g maps from an infinite-dimensional space (the model space \mathcal{M}) to a finite-dimensional one (the data space \mathcal{D}). However one may sidestep this issue by either promoting the data space to infinite dimensions (by assuming that the data are recorded continuously on some or all of the surface of the earth), or by projecting the model space onto the finite-dimensional data space (by assuming that there are the same number of model parameters as data, and assuming the data are independent). In either case the model space and the data space become isomorphic, in which case it is possible that g could be invertible.

The approach of Bleistein and his collaborators^{5 6} is to make a high-frequency Born approximation to g , in which case it can be expressed as a Fourier integral operator (FIO), which can be inverted semi-analytically. There is also a connection between the inversion formulae of Bleistein and the generalized Radon transform discovered by Beylkin.^{4 3} Beylkin derived an approximate inverse (a parametrix in the language of pseudo-differential operators) of the scattering operator in the form of a generalized Radon transform. A clear tutorial on these ideas is provided by Symes.⁴⁴ Using the theory of Fourier integral operators, Symes shows that the output of the adjoint, however computed, has high-frequency strength concentrated in the same places as the parameter perturbations. This is the sense in which the adjoint can be thought of as a migration operator.

In the taxonomy introduced above, direct inversion methods are still imaging methods, however, since they do not allow for an uncertainty analysis. Any attempt to introduce uncertainties inevitably results in replacing Equation 11 with some sort of mis-fit function in order to be able to quantify how well a model fits the data; this puts us right back into the inverse theory soup. Or as B. Efron has said: “Those who ignore Statistics are condemned to reinvent it.” Of course, high-frequency asymptotic methods can be used to approximate g in a waveform inversion procedure. This was done, for example, by Chapman and Orcutt.⁸ A discussion of the connection between FIO inversion methods and classical Kirchhoff migration can be found in the paper by Docherty.¹⁴

0.1.4. Elastic Full-Waveform Inversion

Following Gouveia and Scales²⁰ I will now sketch a field-data example of a nonlinear, elastic waveform inversion calculation based on iteration of Equation 7. The goal is to solve the inference problem for subsurface P- and S-wave impedance and density using surface elastic waveform data and other prior information. There have been a few other applications of Bayesian methods to seismic waveform inversion. For instance, Mora³³ and Crase *et al.*¹¹ compute the maximum *a posteriori* estimate \mathbf{m}_{map} , but without attempting to estimate the covariance matrices C_D and C_M . Thus, they offered a single model as the solution of the inverse problem and do not quantify the data or model uncertainties in the calculation.

The data set considered here consists of vertical-component seismograms from an area where the geology is well described by a horizontally stratified, isotropic elastic parameterization. This allows for rapid, propagator matrix-based synthetic seismogram calculation, which correctly accounts for all or-

ders of multiple scattering and mode conversion, as well as allowing for analytic calculation of the derivative of the forward operator.^{16 18} In addition, well logs (in-situ measurement of elastic moduli and density made in a borehole) are available for the same area and are used to calculate an empirical Bayesian prior for layered earth models. The prior consists both of the background, or *a priori*, model of impedance and density, and the covariance matrix describing the fluctuations about the background. All uncertainties (noise, model and theoretical) are assumed to be described by Gaussian random variables.

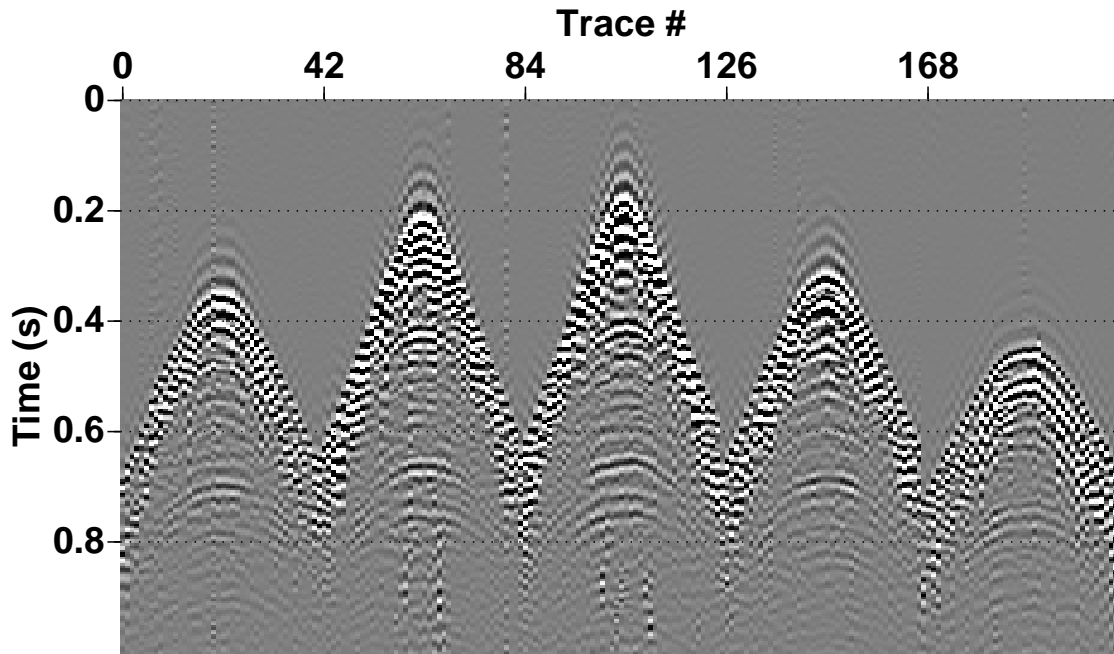
Figure 5 shows several vertical component elastic shot records extracted from a 3D survey provided courtesy of the Reservoir Characterization Project at the Colorado School of Mines. (A shot record is the ensemble of seismograms (or traces) recorded for a given source excitation.) The data were recorded in the Sorrento Basin of Southwestern Colorado, USA. The reservoir rock consists of sediments that filled incised valleys during a series of pulses of rising and falling sea level. The assumption of lateral homogeneity of the subsurface elastic properties is consistent with such a geologic scenario. For a complete description of the results obtained with the Sorrento data, the reader is referred to Gouveia.¹⁸

The borehole measurements (Figure 6) have a resolution on the order of a meter or less, orders of magnitude less than the wavelength of surface seismic waves. In principle, one could incorporate these measurements into the calculation on the same footing as the surface seismic data, but this would have required modeling the borehole measurements. As a simpler alternative, the borehole data are used to compute an empirical Bayesian prior. The idea is this: one does not want to constrain the earth models to be exactly equal to the values obtained by the well log, but simply to reflect the statistical properties of the log. So the goal of this inverse calculation is to find the range of layered earth models that predict the surface seismic data and are close, in a statistical sense, to the well log measurements.

Let the model vector \mathbf{m} be a combination of P- and S-wave impedances and densities, one for each layer in the (presumed) laterally invariant earth: $\mathbf{m} = (\mathbf{m}_{I_p}, \mathbf{m}_{I_s}, \mathbf{m}_\rho)$, where I_p and I_s denote the P- and S-wave impedances, respectively. So if there are m_l layers there will be $3m_l$ total model parameters. (Smooth lateral variations could be incorporated by performing independent inversions at different locations and interpolating the results.)

In order to make the calculations semi-analytic the prior is assumed to be an N -dimensional Gaussian distribution. The mean ($\mathbf{m}_{\text{prior}}$) is estimated

Figure 5 Several vertical component elastic shot records extracted from a 3D seismic survey provided courtesy of the Reservoir Characterization Project at the Colorado School of Mines. The data were recorded in the Sorrento Basin of Southwestern Colorado.



by applying a 200-m running average to the borehole measurements. The covariance is estimated by computing the autocorrelation of the fluctuations of the well log about $\mathbf{m}_{\text{prior}}$ within a sliding window; this accounts for the fact that both mean and variance of the well log change with depth.

Once the correlation $\mathcal{C}(\tau)$ is computed for a given window, the model covariance matrix is estimated by

$$\begin{aligned} C[i, j] &= \mathcal{C}(j - i) \\ &= \frac{1}{N + 1} \sum_{k=i - \frac{N}{2}}^{i + \frac{N}{2}} L(k) L(k + j - i). \end{aligned} \quad (12)$$

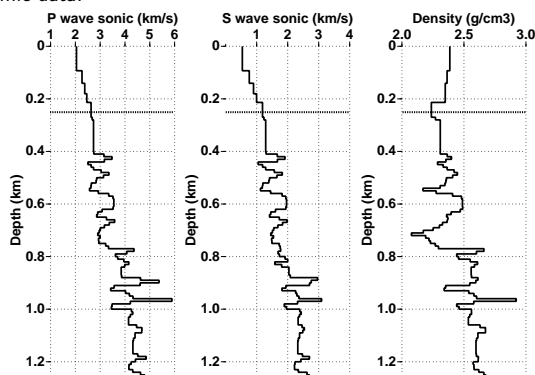
N is the length of the sliding window centered at

depth $i dz$, $L(k)$ is the well-log sample at depth $k dz$, and dz is the depth discretization interval.

Next, the covariance of the data errors must be estimated. In practice this is a difficult task since there are many sources of error, and not all of them can reasonably be assumed to be Gaussian random processes. However, for this calculation a Gaussian fit was made to the four types of data errors regarded as being the most significant: ambient noise, errors in the data processing, uncertainty in the scaling of the synthetic seismograms relative to that of the field data, and discretization errors. Without going into any detail, for each source of noise an ensemble of “noise traces” was generated. Sample covariances were then computed for each ensemble by Monte Carlo simulation. For example, for the ambient noise, pieces of the seismograms were identified that appeared to have no source-generated noise, either very early or very late in the seismograms. Finally all the errors were assumed to be additive, in the sense that the final data covariance matrix is just the sum of the individual covariances.

With these approximations to C_D and C_M , and using the mean of the well logs as the mean of the prior distribution $\mathbf{m}_{\text{prior}}$, one has the full Bayesian posterior. Maximizing this posterior probability is equivalent to minimizing the cost function in Equation 3. A local optimization method (a nonlinear conjugate gradient routine from the free COOL library¹²) was used to find the peak of the posterior; this is the so-called maximum *a posteriori* (MAP) model.

Figure 6 P-wave and S-wave velocities and mass density as functions of depth recorded in a borehole near the location of the surface seismic data.



The nonlinearity of the forward operator makes this problem technically non-Gaussian, so the difference between the mean and the maximum of the posterior might be significant in some problems. In any case, the MAP model is computed by optimizing the cost function, Equation 3.

Once the MAP model is found, the forward operator is linearized about this model. This is equivalent to approximating the posterior by a Gaussian centered about the MAP model. (Note this is not the same thing as linearizing about $\mathbf{m}_{\text{prior}}$, which would make the calculation entirely linear.) Within this approximation the posterior covariance matrix is⁴⁸

$$C_{M'} = [G^T C_D^{-1} G + C_M^{-1}]^{-1}, \quad (13)$$

where G is the linearization of the forward operator about the MAP model.

The MAP models for five particular shots, bracketed by plus or minus one standard-deviation error bars (square roots of the main diagonal of $C_{M'}$) are shown in Figure 7. These error bars do not say anything about the correlation of the parameters; that information is contained in $C_{M'}$ itself (see Gouveia and Scales²⁰ for plots of the posterior covariance).

The ultimate result of this calculation is a probability density on the space of models. This probability takes into account all the probabilistic information that is available, including data uncertainties and *a priori* model uncertainties. This probability defines the range of models that fit the data and are reasonable to a certain degree. One way to distill this information is to look at marginal probabilities. Figure 8 shows marginal probabilities for P-wave impedance at four depths for two different shot records. The dashed lines show the prior marginals derived from the well logs. This is what is known before the surface data have been analyzed. The black lines are the posterior marginals. The inference procedure narrows the marginals to various extents (increases the information); but changes in the means of the pa-

rameters are also evident. From these marginals it is straightforward to answer questions of the form: what is the probability that the P-wave impedance at .6 km depth is between 8.4 and 8.5. (Units for impedances are $10^5 \text{ g s}^{-1} \text{ cm}^{-2}$.)

Figure 7 The Bayes MAP models for five particular shots bracketed by plus or minus one standard deviation error bars. The error bars are derived from the posterior covariance matrix and therefore take into account the prior information as well as information about the forward modeling operator and the various sources of data uncertainty. Rather than thinking of the solution of the problem as consisting of particular models, such as the MAP models, it is better to regard the solution as consisting of these ranges of model parameters; i.e., the posterior probability itself.

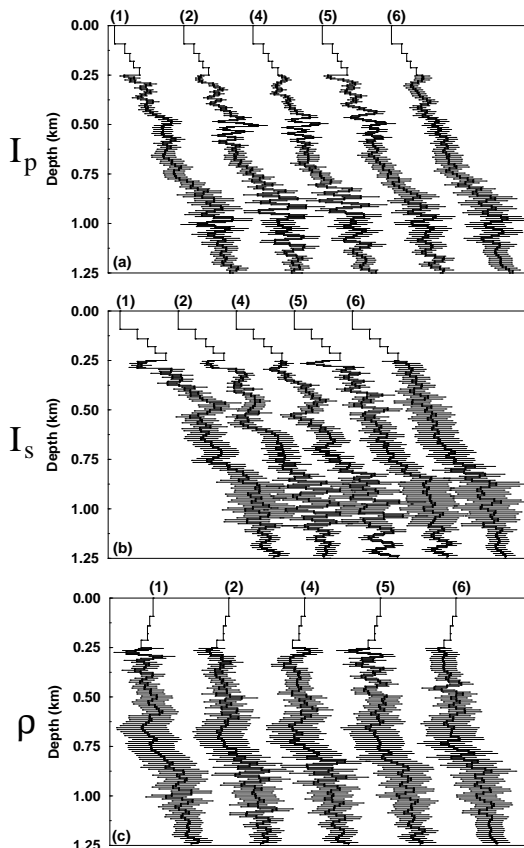
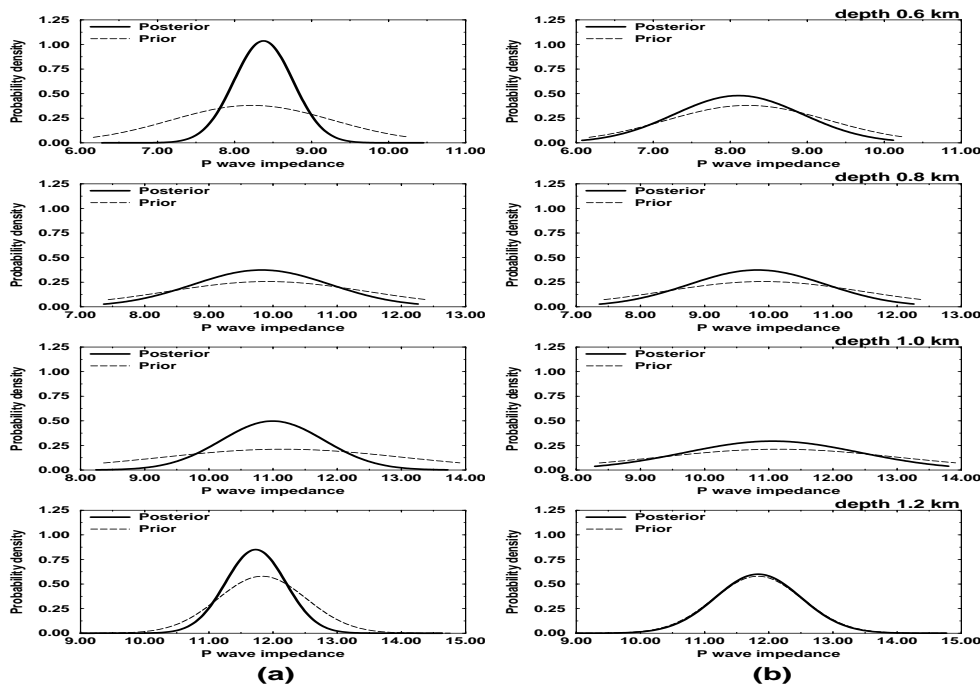


Figure 8 P-wave impedance marginals, at depths 0.6 km, 0.8 km, 1.0 km and 1.2 km. (a) Marginals associated with the data extracted from line 4 of shot gather 1. (b) shows the marginals associated with the data extracted from line 7 of shot gather 5. These two lines represent the range of data quality. The dark lines are the posterior marginals, while the dashed ones are the prior marginals. The ultimate result of the inversion procedure has been to sharpen the marginals, more or less, and to shift the location of the mean values of the parameters.



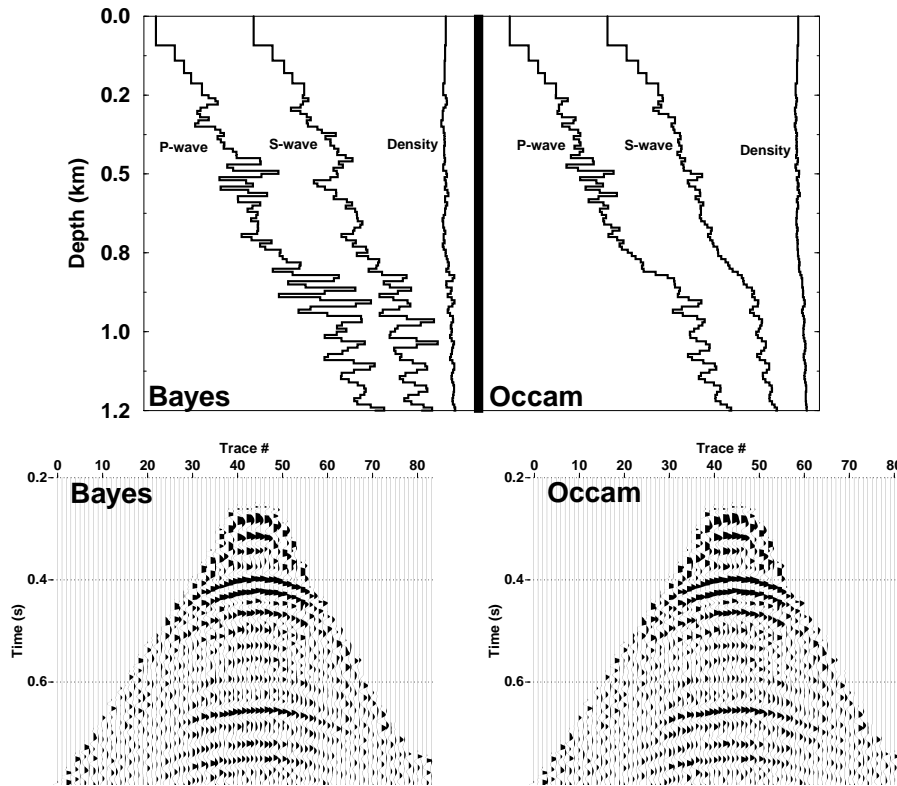
Finally, two different “constructions” from the data are shown in Figure 9. The two models were obtained by local optimization. The model labeled Bayes is one of the MAP models described above. It fits the surface data and is statistically consistent with the borehole measurements. The model labeled Occam is the smoothest model that fits the surface data alone.¹⁹ Below the two models are the data they predict. This should be taken as a cautionary tale—the range of models that is consistent with the surface data is potentially quite large. Imaging alone cannot address this issue. Some quantitative measure of the uncertainty associated with a computed model is necessary. At best, imaging provides a model of the subsurface that may or may not actually explain the data.

In summary, the calculation described here was a nonlinear, elastic waveform inversion calculation applied to field data for which the underlying assumptions of isotropy and lateral invariance appeared to be satisfied. There is no fundamental reason why the same procedure could not be applied to data with anisotropy or complex three-dimensional structure. These may be largely computational considerations since the dominant cost will be the forward modeling. (The calculations described above were performed on

a cluster of low-cost PCs.) On the other hand two key assumptions used here are that the local optimization can be used to find the MAP model starting with the *a priori* model, and the basin of attraction of the MAP model contains most of the support of the *a posteriori* probability distribution. If the former assumption were not true then it might be necessary to perform some sort of Monte Carlo sampling in order to find the basin of attraction of the MAP model. If the latter assumption were not true, then one could not ascribe the same significance to the posterior covariances that has been done here since then the Gaussian approximation would be strongly violated. (What would be the significance of the mean and covariance of a multi-modal probability?) Under such circumstances the inference problem becomes much more complicated theoretically and computationally. For a discussion of Monte Carlo methods applied to Bayesian geophysical inverse problems see the paper by Mosegaard and Tarantola.³⁴

Finally, a philosophical observation. Waveform inversion is sometimes said to be an inherently unstable procedure. But the difference between optimization and inversion should always be kept in mind. The calculation of a particular model, such as the least-squares model, may indeed be ill-conditioned

Figure 9 Two models obtained by local optimization. The model labeled Bayes, is one of the MAP models described above. It fits the surface data and is statistically consistent with the borehole measurements. The model labeled Occam is the smoothest model that fits the surface data alone. Below the two models are the data they predict. In each of the two lower plots, the observed and predicted traces are alternated trace-by-trace for comparison.



or highly nonlinear. But that does not mean that the inversion procedure itself is unstable; it might simply mean that the error bars or confidence sets for some parameters are large. If the uncertainties are too large, then more information or data are required. Particular models, such as the least-squares model, may be of no special significance other than as convenient points in model space about which to center the box of feasible models.

0.1.5. Conclusions

Seismic migration, which is every day applied to vast quantities of exploration seismic data around the world, can be looked upon as a special case of a more general family of waveform inversion algorithms. The goal of migration is to produce maps of the subsurface reflection coefficient. Here, the term “map” is interpreted loosely. It is impossible to make quantitative statements about the accuracy of these maps without performing a full inverse calculation. Migration is used as a form of exploratory data analysis. On the other hand, the goal of seismic waveform inversion is to make quantitative inferences about the earth’s elastic properties. If one

applies a single step of iterative waveform inversion and discards all uncertainty information, then what’s left is a least-squares migration algorithm. But to go from migration to inversion it first is necessary to quantify the uncertainties in data, uncertainties in the data prediction operator, and to characterize objectively the data-independent *a priori* model information. Only then can one analyze the final uncertainties in the model parameters. The difficulty in assessing these uncertainties has heretofore precluded widespread use of waveform inversion methods. I have shown an example where, under reasonable exploration seismic conditions, it is possible to perform elastic full-waveform inversion efficiently with a complete uncertainty analysis—provided certain assumptions about the data uncertainties and the topography of the data mis-fit function are satisfied. An empirical Bayesian procedure was used to integrate surface seismic data with borehole measurements; the latter were instrumental in narrowing the range of feasible models. The final result consists of probabilistic statements about the range of models that fit the surface seismic data and are statistically consistent with the borehole measurements. The techniques used in this inversion procedure could be applied to a wide variety of waveform measurements,

including borehole radar, earthquake source mechanisms, non-destructive testing, ocean acoustics and others.

0.1.6. Acknowledgements

This paper was written while I was on sabbatical in the Laboratoire Ondes et Acoustique of the Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris. I am grateful to my colleagues in Paris for their hospitality during my stay. Thanks to Wences Gouveia for many helpful discussions of the issues related to waveform inversion. Gouveia wrote the code used in the field data example described here. This code, as well as the complete Seismic Unix package, are freely available from the Center for Wave Phenomena's WWW site: <http://www.cwp.mines.edu>. I would also like to thank Ken Larner and Roel Snieder for their reviews. This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, the French Academy of Sciences and the ELF Foundation.

References

1. K. Aki and P. G. Richards. *Quantitative seismology: theory and practice*. Freeman, 1980.
2. K. M. Al-Yahya. Velocity analysis by iterative profile migration. *Geophysics*, 54(06):718–729, 1989.
3. G. Beylkin. The inversion problem and applications of the generalized Radon transform. *Communications in Pure and Applied Mathematics*, 37:579–599, 1984.
4. G. Beylkin. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics*, 26:99–108, 1985.
5. N. Bleistein. On the imaging of reflectors in the earth. *Geophysics*, 52:931–942, 1987.
6. N. Bleistein, J.K. Cohen, and F.G. Hagen. Two and one-half dimensional Born inversion with an arbitrary reference. *Geophysics*, 52:26–36, 1987.
7. R.P. Bording, A. Gersztenkorn, L.R. Lines, J.A. Scales, and S. Treitel. Applications of seismic travel time tomography. *Geophysical Journal of the Royal Astronomical Society*, 90:285–303, 1987.
8. C.H. Chapman and J.A. Orcutt. Least squares fitting of marine seismic refraction data. *Geophysical Journal of the Royal Astronomical Society*, 82:339–374, 1985.
9. G. Chavent and R. Plessix. An optimal true-amplitude least-squares prestack depth-migration operator. *Geophysics*, 64:508–515, 1999.
10. J. Claerbout. *Imaging the Earth's interior*. Blackwell Scientific Publishers, 1986.
11. E. Crase, A. Pica, M. Noble, J. McDonald, and A. Tarantola. Robust elastic nonlinear waveform inversion: Application to real data. *Geophysics*, 55:527–538, 1990.
12. H. Deng, W. Gouveia, and J.A. Scales. The CWP object-oriented optimization library. *The Leading Edge*, 15:365–369, 1996.
13. A. Derode, P. Roux, and M. Fink. Robust acoustic time reversal with high-order multiple scattering. *Physical Review Letters*, 75:4206–4209, 1995.
14. P. C. Docherty. Kirchhoff migration and inversion formulas. *Geophysics*, 56:1164–1169, 1991.
15. M. Fink. Time reversed acoustics. *Physics Today*, 50:34–40, 1997.
16. Fuchs and Mueller. Computation of synthetic seismograms with the reflectivity method and comparison with observations. *Geophys. J.*, 11:417–433, 1971.
17. J. Gazdag. Wave equation migration with the phase-shift method. *Geophysics*, 43:1342–1351, 1978.
18. W. Gouveia. *Bayesian Seismic Waveform Data Inversion: Parameter Estimation and Uncertainty Analysis*. PhD thesis, Colorado School of Mines, 1996.
19. W. Gouveia and J.A. Scales. Resolution in seismic waveform inversion: Bayes vs. Occam. *Inverse Problems*, 13:323–349, 1997.
20. W. Gouveia and J.A. Scales. Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research*, 103:2759–2779, 1998.
21. S. Gray. True-amplitude seismic migration: a comparison of three approaches. *Geophysics*, 62:929–936, 1998.
22. S. H. Gray. Efficient traveltimes calculations for Kirchhoff migration. *Geophysics*, 51:1685–1688, 1986.
23. [Ed] Hale, Dave. *DMO processing*. Soc. Expl. Geophys., 1995.
24. C. Hemon. Equations d'onde et modeles. *Geophysical Prospecting*, 26:790–821, 1978.
25. M. Jannane, W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, D. Trezeguet, and M. Zie. Wavelengths of earth structures that can be resolved from seismic reflection data (short note). *Geophysics*, 54(07):906–910, 1989.
26. S. Jin, R. Madariaga, J. Virieux, and Lambaré. Two dimensional asymptotic iterative elastic inversion. *Geophysical Journal International*, 108:575–588, 1992.
27. R. Kress and R. Spassov. *Integral equation methods in scattering theory*. Wiley, 1984.
28. P. Lailly. The seismic inverse problem as a sequence of before stack migrations. In *Inverse Scattering*. SIAM, 1984.
29. G. Lambaré, Virieux, R. Madariaga, and S. Jin. Iterative asymptotic inversion of seismic profiles in the acoustic approximation. *Geophysics*, 57:1138–1154, 1992.
30. E. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, 1998.
31. D. Loewenthal, L. Lou, R. Robertson, and J. W. Sherwood. The wave equation applied to migration. *Geophysical Prospecting*, 24:85–127, 1976.
32. G. A. McMechan. Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting*, 31:413–420, 1983.
33. P. Mora. Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228, 1987.
34. K. Mosegaard and A. Tarantola. Monte Carlo sampling of solutions to inverse problems. *JGR*, 100:12431–12447, 1995.
35. R.L. Parker. *Geophysical Inverse Theory*. Princeton University Press, 1994.

36. J. A. Scales. Tomographic inversion via the conjugate gradient method. *Geophysics*, 52(02):179–185, 1987.
37. J. A. Scales and R. Snieder. To Bayes or not to Bayes? *Geophysics*, 62(04):1045–1046, 1997.
38. J.A. Scales. Using conjugate gradient to calculate the eigenvalues and singular values of large, sparse matrices. *Geophysical Journal*, 97:179–183, 1989.
39. J.A. Scales. *Theory of Seismic Imaging*. Springer-Verlag, 1995.
40. J.A. Scales and L. Tenorio. Prior information and uncertainty in inverse problems. *In press: Geophysics*, 2000.
41. W. A. Schneider. Integral formulation of migration in two and three dimensions. *Geophysics*, 43:49–76, 1978.
42. P. B. Stark. Minimax confidence intervals in geomagnetism. *Geophys. J. Int.*, 108:329–338, 1992.
43. P.B. Stark. Inference in infinite dimensional inverse problems: discretization and duality. *Journal of Geophysical Research*, 1997.
44. W.W. Symes. Mathematics of reflection seismology. *preprint*, 1999.
45. A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49:1259–1266, 1984.
46. A. Tarantola. Linearized inversion of reflection seismic data. *Geophysical Prospecting*, 32:998–1015, 1984.
47. A. Tarantola. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51:1893–1903, 1986.
48. A. Tarantola. *Inverse Problem Theory*. Elsevier, New York, 1987.
49. J. van Trier and W. W. Symes. Upwind finite-difference calculation of traveltimes. *Geophysics*, 56(06):812–821, 1991.
50. J. C. Vandecar and R. Snieder. Obtaining smooth solutions to large, linear, inverse problems. *Geophysics*, 59(05):818–829, 1994.
51. N. D. Whitmore. Iterative depth migration by backward time propagation. Presented at 53rd Annual International SEG Meeting, Las Vegas, September 1983.
52. Ö. Yilmaz. *Seismic data processing*. SEG, 1987.