

FULL WAVEFIELD INVERSION USING TIME VARYING FILTERS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Patent Application 61/318,561, filed March 29, 2010, entitled FULL WAVEFIELD INVERSION USING TIME
5 VARYING FILTERS, the entirety of which is incorporated by reference herein.

FIELD OF THE INVENTION

[0002] The invention relates generally to the field of numerical inversion of seismic data to infer elastic parameters of the propagating medium. More particularly, the invention is a method for reducing the accuracy requirements on the starting model when performing local
10 objective function optimization in inversions such as seismic data inversion.

BACKGROUND OF THE INVENTION

[0003] Inversion [see, for example, Tarantola, 1984] attempts to find a model that optimally explains observed data. Local inversion methods that minimize the value of an objective function that measures the difference between simulated and observed data, are
15 often the only practical means of solving an inversion problem for a model having a large number of free parameters. These local methods require an initial guess for the model to be inverted. They then iteratively update the model to make it closer to the true solution by searching for a perturbation of the current model in a direction based on the gradient of the objective function. Unfortunately, the objective function often has many minima, not just
20 one minimum corresponding to the solution model. These other minima are called local minima, while the minimum corresponding to the desired solution is termed the global minimum. If the starting model for inversion is too close to the model corresponding to one of these local minima, then local inversion methods will get stuck near that local minimum and never iterate away from it to the global minimum. Thus, the wrong solution is produced
25 no matter how much effort is expended on iteration.

[0004] This local minima problem can be solved by first iterating inversion on an altered objective function that has fewer local minima, but has a global minimum near the location of the desired solution. The result of iterating on this altered objective function should produce a model closer to the desired solution. This more accurate model is then used as the initial

model for iteration on the original objective function. Since this new initial model is close to the global minimum of the original objective function, iteration on the original objective function should now produce an accurate solution. This technique of iterating on an altered objective function is often termed multi-resolution, or multi-grid, or multi-scale inversion, which is discussed further below.

[0005] There are a large number of well known methods of inversion. These methods fall into one of two categories, iterative inversion and non-iterative inversion. The following are definitions of what is commonly meant by each of the two categories:

- Non-iterative inversion – inversion that is accomplished by assuming some simple background model and updating the model based on the input data. This method does not use the updated model as input to another step of inversion. For the case of seismic data these methods are commonly referred to as imaging, migration, diffraction tomography, linear inversion, or Born inversion.
- Iterative inversion – inversion involving repetitious improvement of the subsurface properties model such that a model is found that satisfactorily explains the observed data. If the inversion converges, then the final model will better explain the observed data and will more closely approximate the actual subsurface properties. Iterative inversion usually produces a more accurate model than non-iterative inversion, but is much more expensive to compute.

[0006] Wave inversion means any inversion based on a wave simulator, such as acoustic or seismic inversion. The iterative method most commonly employed in wave inversion is objective function optimization. Objective function optimization involves iterative minimization of the value, with respect to the model $\mathbf{M}$, of an objective function $S(\mathbf{M})$ which is a measure of the misfit between the calculated and observed data (this is also sometimes referred to as the cost function). The calculated data are simulated with a computer programmed to use the physics governing propagation of the source signal in a medium represented by the current model. The simulation computations may be done by any of several numerical methods including but not limited to finite differences, finite elements or ray tracing. Following Tarantola [Tarantola, 1984], the most commonly employed objective function is the least squares objective function:

$$S(\mathbf{M}) = (\mathbf{u}(\mathbf{M}) - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{u}(\mathbf{M}) - \mathbf{d}), \quad (1)$$

where T represent the vector transpose operator and:

$\mathbf{M}$ = the model which is a vector of N parameters $[m_1, m_2, \dots, m_N]^T$,

$\mathbf{d}$ = measured data vector (sampled with respect to source, receiver and time),

5 $\mathbf{u}(\mathbf{M})$ = simulated data vector for model $\mathbf{M}$ (sampled with respect to source, receiver and time),

$\mathbf{C}$ = the covariance matrix.

Objective function optimization methods are either local or global [Fallat, et al., 1999]. Global methods simply involve computing the objective function $S(\mathbf{M})$ for a population of
 10 models $\{\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \dots\}$ and selecting a set of one or more models from that population that approximately minimize $S(\mathbf{M})$. If further improvement is desired this newly selected set of models can then be used as a basis to generate a new population of models that can be again tested relative to the objective function $S(\mathbf{M})$. Global methods are more likely to converge to the correct solution than local methods, but are too expensive to apply to large scale inversion
 15 problems having many model parameters. Well known global inversion methods include Monte Carlo, simulated annealing, genetic and evolution algorithms.

[0007] Local objective function optimization involves:

- | | |
|----|---|
| 20 | <ol style="list-style-type: none"> 1. Set the current model to be the starting model, 2. Compute the gradient, $\nabla_{\mathbf{M}} S(\mathbf{M})$, of the objective function with respect to the parameters that describe the model, 3. Search for an updated model, that is a perturbation of the starting model in a direction based on the gradient, that better explains the observed data. 4. If updated model is not sufficiently accurate then return to step 2 using the updated model as the current model, otherwise terminate. |
|----|---|

25 Algorithm 1 – Algorithm updating a model using local objective function optimization.

Local inversion methods are much more efficient than global methods, and are therefore the only practical methods to apply to a large scale inversion problem. Commonly used local

objective function inversion methods include steepest descent, conjugate gradients and Newton's method.

[0008] It should be noted that computation of $\nabla_{\mathbf{M}} S(\mathbf{M})$, the second step Algorithm 1, requires computation of the derivative of $S(\mathbf{M})$ with respect to each of the N model parameters m_i . When N is very large (roughly more than a thousand), this computation can be extremely time consuming if it had to be performed for each individual model parameter. Fortunately, the adjoint method can be used to efficiently perform this computation for all model parameters at once [Tarantola, 1984]. The adjoint method for the least squares objective function and a gridded model parameterization ($\mathbf{M}$ is a vector with each element representing the model's value in a grid cell) is summarized by the following algorithm:

1. Compute forward simulation of the data using the current model, $\mathbf{M}^{(k)}$ with k being the current iteration, to get $\mathbf{u}(\mathbf{M}^{(k)})$,
2. Subtract the observed data from the simulated data giving $\delta(\mathbf{M}^{(k)})$,
3. Compute the reverse simulation (i.e. backwards in time) using $\mathbf{C}^{-1} \delta(\mathbf{M}^{(k)})$ as the source, producing $\mathbf{u}_{adjoint}(\mathbf{M}^{(k)})$,
4. Finally, $\nabla_{\mathbf{M}} S(\mathbf{M}^{(k)}) = \mathbf{u}_{adjoint}^T(\mathbf{M}^{(k)}) \mathbf{A} \mathbf{u}(\mathbf{M}^{(k)})$, where $\mathbf{A}$ represents the adjoint operator (e.g. identity for gradients with respect to components of $\mathbf{M}$ representing bulk modulus, or spatial gradient for gradients with respect to components of $\mathbf{M}$ representing density).

Algorithm 2 – Algorithm for computing the least-squares cost-function gradient of a gridded model using the adjoint method.

[0009] Local objective function optimization is generally much less expensive than global objective function optimization, but requires a more accurate starting model. This more accurate starting model is required, because the objective function often has many minima and local optimization methods will generally find the closest of these minima. The minimum corresponding to the true model is called the global minimum and all other minima are termed local minima. If the starting model is not nearest to the global minimum then a local optimization technique will likely yield an inaccurate inverted model that corresponds to the closest local minimum. This is illustrated in Fig. 1 where the objective is to invert for

a model **M** which has two parameters m_1 and m_2 . The dashed contours **110** display the values of the objective function as a function of the parameters m_1 and m_2 . The global minimum **120** is marked by a solid black circle and two local minima **130** and **140** are shown by gray filled circles. The inversion starts at initial model $\mathbf{M}^{(0)}$ (**150**) and proceeds by local optimization to the iteration one model $\mathbf{M}^{(1)}$ and so forth to model $\mathbf{M}^{(3)}$ (**160**). No matter how many more iterations of local optimization are attempted the inverted model will only get closer to the local minimum **130** near $\mathbf{M}^{(3)}$, rather than approximating the global minimum **120**.

[0010] Several methods have been proposed that attempt to overcome this local minima problem. As mentioned above, many of these methods involve iterating on an altered objective function during the early iterations of the inversion. This altered objective function is chosen to have fewer local minima, but to have a global minimum near the original objective function's global minimum. By this means, the early iterations will produce a model that though inaccurate, is significantly closer to the original objective function's global minimum. Figure 2 illustrates a local optimization corresponding to Fig.1 but using an altered objective function having fewer local minima. The altered objective function has a global minimum **210** (the solid black circle) that is close to but not at the same location as the global minimum **220** for the original objective function (the cross-hatched circle). Starting from the initial model $\mathbf{M}^{(0)}$ (**230**), which is the same as the initial model used in Fig. 1, two iterations using the altered objective function results in a model $\mathbf{M}^{(2)}$ (**240**). This model $\mathbf{M}^{(2)}$ can then be used as the initial model for further iterations but now using the original objective function. This is illustrated in Fig. 3 where the iteration 2 model from Fig. 2 (shown in gray), renumbered **310**, is used as the starting model. Iteration now converges to a model $\mathbf{M}^{(4)}$ (**320**) near the global minimum **220** rather than near a local minimum as in Fig. 1. Because the starting model is more accurate than the original starting model, the inversion now iterates to the correct solution.

[0011] Typically when altering the original objective function, the number of local minima in the altered objective function is inversely related to the distance between the global minima of the original and that of the altered objective function. Thus, it can be advantageous to iterate on a sequence of altered objective functions starting with one having the fewest number of local minima and least accurate global minimum, proceeding through objective functions that have increasing numbers of local minima and increasing accuracy of the global minim, then ending by iterating on the original objective function. Methods that

perform initial iterations on altered objective functions having few local minima are often referred to as multi-scale or multi-grid methods, and a flow chart for this technique is illustrated in Figure 4.

[0012] The process starts at step 410 by choosing an alteration of the original objective function to optimize. This altered objective function, which depends on the data to be fit 420, is iterated at step 430 until the altered objective function is found to be sufficiently minimized at step 440. (The value is less than a selected maximum or another stopping condition is met.) When that occurs, it is determined at step 450 whether the current inverted model sufficiently minimizes the original objective function. If not, the process returns to step 410 and either choose a new altered objective function or the original objective function to optimize. Eventually, the process ends (460).

[0013] Two altered objective functions have been proposed in the literature for solving the local minima problem in seismic full wave field inversion ("FWI"):

- 15 • **High cut filters** - Bunks (Bunks et al., 1995) describes altering the least squares objective function by applying time invariant high-cut filters (sometimes called a low-pass filter, meaning a filter that passes frequencies below its cutoff frequency and rejects frequencies above it) to both the measured data and the source signature used for computing the simulated seismic data. The high-cut, i.e. cutoff, frequency of these filters is then increased as the inversion is iterated, with no filter being applied for the final iterations (no filter corresponding to the original objective function). It is well known how to design such filters; see, for example, Press et al., *Numerical Recipes in FORTRAN, The Art of Scientific Computing*, Cambridge University Press (1992). They may also be obtained from sources such as Seismic Un*x (see <http://www.cwp.mines.edu/cwpcodes/>).
- 25 • **Layer stripping** – Maharramov (Maharramov et al., 2007) teaches that initial iterations of inversion should be localized to shallow layers, and that this depth range be extended as iteration proceeds. Correspondingly, when only shallow depths are inverted, then only shorter times in the data are inverted, because a shallow model can only predict the shorter time portions of the data.

In general FWI will converge to the global minimum if the starting model is accurate enough to predict traveltimes for any propagation mode to within a half period of that mode. This may be called Pratt's criterion. ("In practice, with seismic waveform inversion this implies that much of the waveform energy must be predicted (by the initial model) to within a half-wavelength of the observed waveforms; if not, a minimum misfit model will be obtained when the predicted waveforms match the wrong cycle in the observed waveforms." -- Pratt, 1999).

[0014] The present invention is an improved method for reducing the accuracy requirements on the starting model when performing local objective function optimization.

10

SUMMARY OF THE INVENTION

[0015] The present inventive method is applicable to any inversion based on a wave simulator, such as acoustic or seismic inversion. In one of its aspects, the invention is a specific method for altering the objective function which will, for a given inaccuracy in the starting model, reduce the number of iterations needed to find the global minimum. Reducing the number of iterations will correspondingly reduce cost and compute time. The alteration comprises incorporating a time varying filter into the objective function. This filter is chosen so that some statistical measure of the traveltime difference between the measured and computed data is less than some fraction (usually one quarter) of the dominant period of the data. This implies that the filter be a high-cut filter. The filter is further chosen such that the high-cut frequency of this filter decreases with increasing traveltime, making it a *time-varying* filter.

[0016] With reference to the flow chart of Fig. 6, in one embodiment the invention is a method for inverting measured seismic data (65) to infer a physical property model of a subsurface region, comprising successively updating the model by performing iterative, multi-scale inversion of the measured seismic data using local-minimum optimization (64) of an objective function that computes misfit between model-simulated seismic data and the measured seismic data (61), wherein a changing low-pass filter is used to successively alter the objective function from scale to scale by filtering the measured and simulated seismic data in the misfit computation (62), said filter being time-varying wherein the filter's low-pass cutoff frequency varies with traveltime of the seismic data being filtered at one or more scales of the inversion. The filter may be designed using Pratt's criterion for reducing the

number of local minima, modified to involve a statistical measure (such as the mean) of the data with respect to source and receiver location but not with respect to traveltimes, leaving the filter time-varying (63). The final stage of the multi-scale inversion, preferably using the unaltered objective function, is thus more efficiently provided with a more accurate starting
5 model to help it converge to a global minimum, resulting in an optimized physical properties model (66).

[0017] As with any data inversion, the process in practical applications is highly automated, i.e. is performed with the aid of a computer programmed in accordance with the disclosures herein.

10

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The present invention and its advantages will be better understood by referring to the following detailed description and the attached drawings in which:

Fig. 1 is a schematic illustration of an inversion converging to a local minimum;

Fig. 2 is a schematic diagram of local optimization corresponding to Fig. 1 but using
15 an altered objective function having fewer local minima;

Fig. 3 illustrates local optimization with the original objective function from Fig. 1 but using the iteration 2 model from Fig. 2 (shown in gray) as the starting model;

Fig. 4 is a flow chart showing basic steps in multi-scale optimization;

Fig. 5 is a schematic diagram illustrating the traveltimes error, t_{error} , between the
20 measured and simulated data, and the instantaneous period, T , of a seismic trace; and

Fig. 6 is a flow chart showing basic steps in one embodiment of the present invention.

[0019] The invention will be described in connection with example embodiments. However, to the extent that the following detailed description is specific to a particular embodiment or a particular use of the invention, this is intended to be illustrative only, and is
25 not to be construed as limiting the scope of the invention. On the contrary, it is intended to cover all alternatives, modifications and equivalents that may be included within the scope of the invention, as defined by the appended claims.

DETAILED DESCRIPTION OF EXAMPLE EMBODIMENTS

[0020] Mathematically, Pratt's criterion can be stated as:

$$\frac{\max_{s,r,t} |t_{error}(\mathbf{M}, s, r, t)|}{\max_{s,r,t} T(s, r, t)} \leq \frac{1}{2}, \quad (2)$$

where t_{error} is the traveltime error between the measured and simulated data, and T is the instantaneous period of the measured data as illustrated in Fig. 5. In Fig. 5, the traveltime error between the measured and simulated data is indicated as t_{error} , and the instantaneous period of a seismic trace is indicated by T . The traveltime error is the amount of time shift needed align the measured and simulated data. The instantaneous period is the time between similar phases (e.g. peaks or troughs) of the data. (Traveltime means the elapsed time from generation of a seismic wave by the seismic source until the wave is recorded at a receiver.)

[0021] Both the traveltime error and instantaneous period are functions of the source s , the receiver r and the traveltime t . In addition t_{error} is a function of the accuracy of the current model $\mathbf{M}$. In practice Equation 2 may be more restrictive than necessary to ensure convergence of FWI. In particular a less rigid statistical measure than the maximum (e.g., the mean) could be used or a target other than 1/2 could be used on the right side of the inequality. Thus, in practice, Equation 2 can be replaced by the following:

$$\frac{stat_{s,r,t} |t_{error}(\mathbf{M}, s, r, t)|}{stat_{s,r,t} T(s, r, t)} \leq \frac{\alpha}{2}, \quad (3)$$

where $stat$ is some statistic such as the mean, mode or mean squared error, and α is a constant that is approximately equal to one. It is not expected that the results will be very sensitive to the choice of statistic.

[0022] Bunks's alteration of the objective function follows a line of reasoning similar to Pratt's criterion by using a high-cut filter to increase $T(s, r, t)$ thus allowing larger values of t_{error} . It is understood that a main cause of local minima is cycle skipping, and longer periods make this less likely. In theory, t_{error} could be reduced instead of limiting the data to lower frequency; however the only way to do this would be to have a more accurate starting model, which is very difficult and maybe impossible. Furthermore the goal of FWI is to produce a

more accurate model, so requiring a very accurate starting model reduces the value of FWI. On the other hand Maharramov's layer stripping method avoids large traveltimes errors by inverting only a shallow model that propagates only modes that have small traveltimes. Since traveltimes errors usually increase with traveltimes, limiting the inversion to shorter traveltimes keeps t_{error} within the rule of thumb.

[0023] In the present invention, an alternative is proposed to Equation 3's "rule of thumb" that will lead to a new strategy for ensuring convergence. The alternative rule of thumb is as follows:

$$\frac{\underset{s,r}{stat} |t_{error}(\mathbf{M}, s, r, t)|}{\underset{s,r}{stat} T(s, r, t)} \leq \frac{\alpha}{2}. \quad (4)$$

This rule of thumb differs from Equation 3 in that the statistical analysis is no longer performed with respect to time. After applying the statistical calculations, the numerator and denominator on the left side of Equation 4 are not functions of the source and receiver locations. Equation 4 is equivalent to:

$$\hat{T}(t) \geq \frac{2}{\alpha} \hat{t}_{error}(\mathbf{M}, t), \quad (5)$$

where $\hat{T}(t) = \underset{s,r}{stat} T(s, r, t)$, and $\hat{t}_{error}(\mathbf{M}, t) = \underset{s,r}{stat} |t_{error}(\mathbf{M}, s, r, t)|$.

[0024] In practice $\hat{t}_{error}$ is an increasing function of traveltimes t , because traveltimes errors tend to accumulate as waves propagate through an inaccurate model. Equation 5 suggests that the optimal strategy for multi-scale inversion would be to alter the measured and simulated data so that the average instantaneous period of the seismic data increases with traveltimes in a manner similar to $\hat{t}_{error}$. This can be accomplished by applying a time-varying high-cut filter to the measured data. The high-cut frequency of this time varying filter should decrease with increasing traveltimes. The advantage of this approach over the Bunks technique is that more information (i.e. higher frequency information at small traveltimes) would be used for the early iterations of inversion, thus better constraining the inverted model leading to faster convergence. The advantage of our proposal relative to Maharramov's layer stripping is again that more of the data (i.e. seismic modes propagating

through deeper portions of the model) would be used in early iterations, leading to faster convergence.

[0025] The function $\hat{t}_{error}$ depends on both how traveltime errors are measured and also on what statistical measure is applied to these traveltime error measurements. However, it
 5 may be expected that this proposed multi-scale inversion strategy will not be strongly sensitive to $\hat{t}_{error}$. In fact, rather than put significant effort into measuring $\hat{t}_{error}$ from the data, an alternative strategy would be to assume a simple functional form for $\hat{t}_{error}$, such as the linear function $\hat{t}_{error} = \beta(\mathbf{M}_0)t$ where $\mathbf{M}_0$ is the initial model. This assumed functional form would then be used to design a time-varying high-cut filter satisfying Equation 5, and
 10 inversion using this filter would be attempted. If the inversion does not converge then the value of β would be increased and inversion with this more conservative estimate of $\hat{t}_{error}$ would be attempted. This testing would continue until a β is found that yields convergence for the initial model.

[0026] After finding a β that converges for the initial model, $\mathbf{M}_0$, then iteration will
 15 produce a current inverted model which is more accurate than the initial model. This increased accuracy implies that β should be reduced as iteration proceeds. This decrease in β implies a corresponding time-varying filter that passes higher frequencies. The inversion proceeds with time-varying filters that pass more and more high frequencies, until all frequencies in the data are used in the final iterations of inversion.

[0027] To make practical use of time-varying filters in inversion, it is necessary to be able
 20 to compute the adjoint gradient (Algorithm 2) in a manner consistent with the time varying filter. The mathematically most straight forward way of doing this is to implement the time varying filter using the inverse covariance matrix $\mathbf{C}^{-1}$ in Equation 1. To do this, the inverse covariance matrix $\mathbf{C}^{-1}$ would be chosen to be non-diagonal (in the time dimension), with
 25 elements equal to a temporal representation of the time varying filter coefficients. Since the filter would be time varying, the filter coefficients vary with time. Example 1 shows an

example sub-matrix of C^{-1} , corresponding to a particular source and receiver, that implements a time-varying filter. The first row of this sub-matrix is zero, excepting a one on the diagonal.

$$\begin{array}{c}
 \text{time} \rightarrow \\
 \begin{pmatrix}
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & .95 & .05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & .90 & .10 & .05 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & .85 & .15 & .10 & .05 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & .80 & .20 & .15 & .10 & .05 & 0 \\
 0 & 0 & 0 & 0 & 0 & .75 & .25 & .20 & .15 & .10 \\
 0 & 0 & 0 & 0 & 0 & 0 & .70 & .30 & .25 & .20 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & .65 & .35 & .30 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .60 & .40 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .55
 \end{pmatrix}
 \end{array}
 \begin{array}{c}
 \text{time} \downarrow
 \end{array}$$

5

Example 1

This implies that this particular time varying filter performs no filtering at time zero. The off-diagonal elements increase for increasing rows in the sub-matrix, implying that the high cut frequency of this time varying filter decreases with increasing time. Notice that the traveltimes error could be viewed as functions of the source or receiver. In this case, this method could be applied in a more general manner than just a simple time varying filter. For example:

1. The traveltimes error is often not just a function of traveltimes, but is also often a function of the offset between the source and the receiver. If this is the case, it would be advantageous to use different time-varying filters for different source receiver offsets. As iteration proceeds the filter would be relaxed as the accuracy of the model increases.
2. The traveltimes error is often a function of the time dip of the data in the source and/or receiver domain. This can happen because steep time dips correspond to waves traveling in predominantly horizontal direction which are more sensitive to the accuracy of the initial model. In this case, time-varying dip-filters (e.g. a frequency-wavenumber filter that removes seismic events having high time dip) could be used in place of time-varying temporal-filters. As

20

iteration proceeds the dip-filter would be relaxed as the accuracy of the model increases.

[0028] In any case the filter (e.g. a space- and time-varying filter, a time-varying dip-filter etc.), can be implemented in the covariance matrix $\mathbf{C}^{-1}$ as explained above, and the
5 computation of the gradient then also proceeds as described above.

[0029] A preferred approach might be that if $\hat{t}_{error}(\mathbf{M}, t)$ can be estimated for the current model $\mathbf{M}$, then the time varying filter should be designed to be consistent with Equation 5. Otherwise, it is reasonable to estimate that $\hat{t}_{error}(\mathbf{M}, t)$ is a linear function βt , using an initial guess for the value of β . Again the time varying filter should be designed to be consistent
10 with both this estimate of β and Equation 5. If this inversion converges then the process is completed. If the inversion does not converge, then β would be increased and another inversion would be attempted. This process of increasing β would continue until convergence is achieved.

[0030] In practice the matrix $\mathbf{C}^{-1}$ representing the time varying filter would be very
15 sparse, and therefore its application to the data residual in step 3 of algorithm 2 would best be accomplished by application of a time varying filter operator rather than by matrix multiplication. In fact this method of inversion is probably not strongly sensitive to the method used to implement the time varying filter. Therefore, time varying filter could be most efficiently implemented as a windowed time invariant filter. In other words, the data
20 would be separated into time windows, and then different time invariant filters would be applied to the different windows.

[0031] The foregoing description is directed to particular embodiments of the present invention for the purpose of illustrating it. It will be apparent, however, to one skilled in the art, that many modifications and variations to the embodiments described herein are possible.
25 For example, the inventive method is not limited to seismic data, and can be applied to any data where multiscale inversion is used to avoid local minima. All such modifications and variations are intended to be within the scope of the present invention, as defined by the appended claims.

References

- Bunks, C., F. M. Saleck, S. Zaleski, G. Chaven, 1995, "Multiscale seismic waveform inversion," *Geophysics*, **60**, pp. 1457-1473.
- Maharramov, M., U. Albertin, 2007, "Localized image-difference wave-equation tomography," SEG Annual Meeting Expanded Abstracts, San Antonio, 2007, pp. 3009-3013.
5
- Fallat, M. R., Dosso, S. E., "Geoacoustic inversion via local, global, and hybrid algorithms," *Journal of the Acoustical Society of America* **105**, pp. 219-3230 (1999).
- Pratt, R. G., "Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model," *Geophysics* **64**, pp. 888-901 (1999).
- 10 Tarantola, A., "Inversion of seismic reflection data in the acoustic approximation," *Geophysics* **49**, pp. 1259-1266 (1984).

CLAIMS

1. A method for inverting measured seismic data to infer a physical property model of a subsurface region, comprising successively updating the model by performing iterative, multi-scale inversion of the measured seismic data on a computer using local-minimum optimization of an objective function that computes misfit between model-simulated seismic data and the measured seismic data, wherein a changing low-pass filter, hereinafter called the filter, is used to successively alter the objective function from scale to scale by filtering the measured and simulated seismic data in the misfit computation, said filter being time-varying wherein the filter's low-pass cutoff frequency varies with travelttime of the seismic data being filtered at one or more scales of the inversion.
2. The method of claim 1, wherein the filter's low-pass cutoff frequency decreases with increasing travelttime.
3. The method of claim 1, wherein each successively altered objective function corresponds to the filter being changed to reject less data.
4. The method of claim 3, wherein the objective function is successively altered until a final scale is reached at which the objective function is unaltered and wherein iteration continues until a convergence criterion is satisfied or other stopping condition is met.
5. The method of claim 4, wherein the filter corresponding to the final scale passes all data and rejects none.
6. The method of claim 3, wherein the filter is changed to reject less data by increasing the filter's low-pass cutoff frequency.
7. The method of claim 1, wherein one or more iterations are performed at each scale of the multi-scale inversion, i.e. for each change of the high-pass filter and alteration of the objective function.

8. The method of claim 1, wherein the local-minimum optimization of the objective function comprises computing the gradient of the objective function with respect to one or more parameters of a current physical property model, then searching for an updated physical property model, that is a perturbation of the current physical property model in a direction based on the gradient that better explains the measured seismic data.
9. The method of claim 8, wherein the gradient is computed by an adjoint method.
10. The method of claim 1, wherein the inversion is full wavefield inversion.
11. The method of claim 2, wherein to ensure that the iterations converge to a global minimum of the objective function, the following criterion is used to design a time-varying low-pass filter: the model when iteration begins should be accurate enough to predict traveltimes for any propagation mode to within a half period of that mode after the time varying filter is applied.
12. The method of claim 11, wherein said convergence criterion may be mathematically expressed as:

$$\frac{\text{stat}_{s,r} |t_{\text{error}}(\mathbf{M}, s, r, t)|}{\text{stat}_{s,r} T(s, r, t)} \leq \frac{\alpha}{2}$$

where $\mathbf{M}$ represents the physical property model, t_{error} is traveltime error between the measured and simulated seismic data, T is the measured data's instantaneous period, α is a selected constant, t is seismic wave traveltime, s represents seismic source coordinates, r represents seismic receiver coordinates, and stat means mean or average or another measure to reduce variability of s and r to constants.

13. The method of claim 12, wherein α is a number within the range $\frac{1}{2}$ to 1.

14. The method of claim 12, where rate of decrease of the low-pass cutoff frequency with increasing travelttime is determined consistent with the convergence criterion.
15. The method of claim 1, further comprising using different time-varying filters for different source-receiver offsets.
16. The method of claim 9, wherein the filter is implemented using the inverse covariance matrix $\mathbf{C}^{-1}$ in Equation 1.
17. The method of claim 16, wherein the inverse covariance matrix $\mathbf{C}^{-1}$ is chosen to be non-diagonal in the time dimension, with elements equal to a temporal representation of the time varying filter's coefficients, said coefficients varying with time.
18. The method of claim 12, wherein $\hat{t}_{error} = stat_{s,r} \left| t_{error}(M, s, r, t) \right|$ is assumed to be a linear function of travelttime of the form $\hat{t}_{error} = \beta(\mathbf{M}_0)t$, where $\mathbf{M}_0$ is an initial choice for the model when iteration begins, and β is a selected constant, wherein if the iterative inversion does not converge using this filter design, then the value of β would be increased and inversion with this more conservative estimate of $\hat{t}_{error}$ would be attempted, and so on until a β is found that yields convergence for the initial model.
19. The method of claim 1, wherein the inversion is full wavefield inversion.
20. The method of claim 1, wherein a computer is programmed to perform at least some of the method's steps.
21. A computer program product, comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement a method for full wavefield inversion of measured data to infer a model of a propagating medium for the wavefield, said method comprising:

successively updating the model by performing iterative, multi-scale inversion of the measured data using local-minimum optimization of an objective function that computes misfit between model-simulated data and the measured data;

wherein a changing low-pass filter, hereinafter called the filter, is used to successively alter the objective function from scale to scale by filtering the measured and simulated data in the misfit computation, said filter being time-varying;

wherein the filter's low-pass cutoff frequency varies with wavefield traveltime through the medium corresponding to the data being filtered at one or more scales of the inversion.

22. The computer program product of claim 21, wherein the filter's low-pass cutoff frequency decreases with increasing traveltime.

23. The computer program product of claim 21, wherein the data being inverted are seismic data, and the propagating medium is a subsurface region of the earth.