CWP-183
May 1995

Using multi-resolution analysis to study the complexity of inverse calculations

H. Lydia Deng

Using multi-resolution analysis to study the complexity of inverse calculations

H. Lydia Deng

ABSTRACT

Optimization is a tool for many inverse calculations. However, in practice it is found that functions we wish to optimize are often highly non-convex (multi-modal). An example of optimizations with such difficulties is in seismic waveform inversion. The waveform misfit functions are generally multi-modal, partly because of the oscillatory nature of seismic waves.

In this paper, I show how multiresolution analysis can be used to deal with the multi-modal nature of objective functions arising in seismic inversion problems. Using residual statics estimation as an example, this paper shows that the waveform misfit function is multi-modal even for the simplest case, and, indeed, MRA simplifies the waveform misfit function. By studying the complexity of objective functions on MRA decomposed data, increased understanding of the complexity of objective functions, such as those arising in inverse problems, can be gained.

REALITY IN OPTIMIZATIONS

Optimization is a tool for many inverse calculations. However, in practice it is found that functions we wish to optimize, *objective functions*, are often highly non-convex (multi-modal). Many optimization algorithms have been developed to handle different inverse problems; each of them usually works well for some situations but fails for the others. Therefore, it is important to understand what makes some inverse problems difficult, while others are not.

The multi-modalities in optimizations are especially serious for seismic waveform misfit functions. This is partly caused by the oscillatory nature of seismic waves. When there are many local minima, the gradient-based searching methods have little chance of finding the correct global minimum.

Global search methods are used by many researchers when dealing with multi-modal misfit functions. Rothman (1985; 1986) solved a residual statics problem by simulated annealing (SA). Scales et al. (1992), Smith et al. (1992), Sen and Stoffa (1991a; 1991b; 1992) and Gouveia (1993) studied SA and genetic algorithms (GAs) on a variety of multi-modal optimization problems, received satisfying results for problems they studied.

Although it can be proved that SA and GAs converge to global extrema asymptotically, it is not guaranteed that they would find the global extrema in a finite amount of computational time. Gouveia (1994) studied some hybrid methods of the traditional gradient-based searching and distributed-parallel GAs. They show that these hybrid methods are more efficient for the residual statics estimation than a distributed GA.

Many researchers found other alternative ways to overcome problems of local minima, while keeping the computation effort relatively low. Here I describe two strategies.

The first approach is to choose initial models that are close to the global extrema by integrating into the solution formalism *a priori* knowledge other than data itself. Although applying *a priori* information to inversions does not reduce the complexity of the objective function in global sense, it helps to confine the searches to a smaller range, which increases the chance of convergence to global minima (Tarantola, 1987). In a seismic waveform inversion, Chapman (1985; 1988) suggested using travel-time information to infer a smooth velocity model, which is used as the initial guess to the waveform inversion. Scales and Tarantola (1994) conducted statistical analysis on geologic and well-log information in order to obtain *a priori* information for waveform inversion.

The second approach is to simplify the objective function of optimizations so as to reduce the total number of local minima. Shaw and Orcutt (1985) suggested using the envelope of seismic data as the fitting target for simplifying the objective function of waveform inversions. Unfortunately, they found that the "envelope" in seismic data was sensitive to noise.

Despite the failure of "envelope inversion", there has been significant interest in simplifying the waveform misfit function while not having to extract indirect information from the data. In waveform inversions, the success of the differential semblance optimization (DSO) method (Symes and Carazzone, 1991; Symes, 1993) demonstrates that the complexity of objective functions in inversions is significantly affected by the parameterization of the target models. Chevent (1994) and Symes (1994) proved theoretically that DSO produces almost convex objective functions in some waveform inversions.

Multi-scale ideas are also used to deal with the multi-modality of objective functions in inverse calculations. Seismic waveform data can be decomposed into several data sets by low-pass filters, each of which contains progressively higher frequency data. Optimizations are applied to these data sets iteratively in order to increase the chance of finding the global minima (Saleck et al., 1993; Chen, 1994). This paper describes a similar approach by means of the multiresolution analysis (MRA). As the first step of studying the complexity of inverse problems, this paper studies a simple residual statics problem.

MULTIRESOLUTION ANALYSIS

What is a Multiresolution Analysis?

Multiresolution analysis (MRA) was formulated based on the study of orthonormal, compactly supported wavelet bases. Wavelets theory and its applications are rapidly developing fields in applied mathematics and signal analysis. Wavelet basis representation of certain signals show advantages over the traditional Fourier basis representation both theoretically and practically. The MRA concept was initiated by Meyer (1992) and Mallat (1989), which provides a natural framework for the understanding of wavelet bases. Here, I give a brief description of orthonormal, compactly supported wavelet bases; detailed information can be found, for example, in Daubechies (1992) and Jawerth and Sweldens (1994).

An orthonormal, compactly supported wavelet basis of $\mathbf{L}^2(\mathbf{R})$ is formed by the dilation and translation of a single function $\psi(x)$, called the wavelet function:

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k); \quad j,k \in \mathbf{Z}, \tag{1}$$

where $\mathbf{Z}$ is the set of integers. In equation (1), the function ψ has M vanishing moments up to order $M-1$, and it satisfies the following "two-scale" difference equation,

$$\psi(x) = \sqrt{2}\sum_{k=0}^{L-1} g_k\psi(2x-k). \tag{2}$$

The wavelet function $\psi(x)$ has a companion, the scaling function $\phi(x)$, which also forms a set of orthonormal bases of $\mathbf{L}^2(\mathbf{R})$,

$$\phi_{j,k}(x) = 2^{-j/2}\phi(2^{-j}x - k); \quad j,k \in \mathbf{Z}. \tag{3}$$

The scaling function $\phi(x)$ satisfies,

$$\int_{-\infty}^{+\infty} \phi(x)dx = 1.$$

and the "two-scale difference" equation,

$$\phi(x) = \sqrt{2}\sum_{k=0}^{L-1} h_k\phi(2x-k). \tag{4}$$

In equations (2) and (4), two coefficient sets $\{g_k\}$ and $\{h_k\}$ have the same finite length L for a certain basis, where L is related to the number of vanishing moments M in $\psi(x)$. For example, L equals $2M$ in the Daubechies wavelets. In the wavelet representation of signals, $\{h_k\}_{k=0,\dots,L-1}$ behaves as a low-pass filter and $\{g_k\}_{k=0,\dots,L-1}$ behaves as a high-pass filter to signals. These two filters are related by

$$g_k = (-1)^k h_{L-k}; \quad k = 0, \dots, L-1, \tag{5}$$

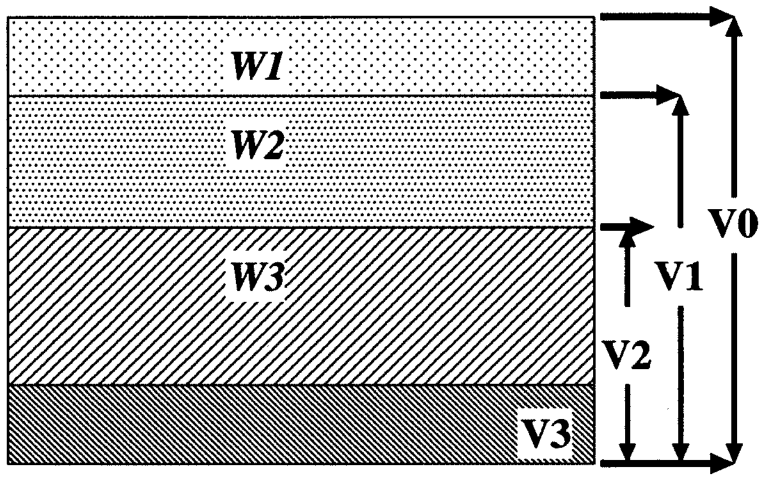

FIG. 1. Illustration of the sequence of multiresolution analysis subspaces V_j. W_j is the orthogonal complement of V_j in V_{j-1}. Space V_0 represents the space that contains the finest resolution data, and $V_0 = V_3 \oplus W_3 \oplus W_2 \oplus W_1$.

and are called *quadrature mirror filters* (QMF). An extensive study of the QMF can be found in (Monzon, 1994).

The MRA of $\mathbf{L}^2(\mathbf{R})$ is a set of nested, closed subspaces $\{V_j;\ j \in \mathbf{Z}\}$, such that

$$... V_3 \subset V_2 \subset V_1 \subset V_0 ..., \tag{6}$$

where the basis for the subspace V_j is a set of orthonormal, translated functions, and each of these functions sets is a fixed dilation of the scaling function, $\{\phi_{j,k};\ k \in \mathbf{Z}\}$. Therefore, these subspaces have the property

$$f(x) \in V_0 \Longleftrightarrow f(2^{-j}x) \in V_j; \quad \forall j \in \mathbf{Z}. \tag{7}$$

Defining W_j to be the orthogonal complement of V_j in V_{j-1}, they are related by

$$V_{j-1} = V_j \oplus W_j. \tag{8}$$

The wavelet basis $\{\psi_{j,k}; k \in \mathbf{Z}\}$, as in equations (1) and (2), forms the orthonormal basis of the subspace W_j. Therefore, for $j < n_0$, we can have

$$V_j = V_{n_0} \oplus W_{n_0} \oplus W_{n_0-1} \cdots \oplus W_{j+1}. \tag{9}$$

Figure 1 illustrates the nesting of subspaces V_j and their orthogonal complements W_j. In Figure 1, V_0 contains the original data which has the finest resolution; the projection of the data on $\{V_j;\ j = 1, 2, 3\}$ has increasingly coarser resolution. In this paper, the data projected onto the subspace V_j is referred as the decomposition of data at resolution level j.

We define the projection of a function $f \in V_0$ on V_j to be $f^j(x)$. Then the jth resolution level of the function has the form

$$f^j(x) = \sum_k s_{j,k}\, \phi_{j,k}(x), \tag{10}$$

where $s_{j,k}$ is the projection of the function $f(x)$ on the basis $\phi_{j,k}$; that is,

$$s_{j,k} = \int f(x)\, \phi_{j,k}(x)\, dx.$$

Next, define the projection of $f(x)$ on the subspace W_j to be

$$df^j(x) = \sum_k d_{j,k}\, \psi_{j,k}(x), \tag{11}$$

where $d_{j,k}$ is the projection of function $f(x)$ on the basis $\psi_{j,k}$

$$d_{j,k} = \int f(x)\, \psi_{j,k}(x)\, dx.$$

Then, equation (9) implies that the original function $f(x) \in V_0$ can be represented by

$$\begin{aligned} f(x) &= f^{n_0}(x) + \sum_{j=n_0}^{1} df^j(x) \\ &= \sum_k s_{n_0,k}\, \phi_{n_0,k}(x) + \sum_{j=n_0}^{1} \sum_k d_{j,k}\, \psi_{j,k}(x). \end{aligned} \tag{12}$$

Figure 2 shows the decomposition of a simple synthetic seismic trace at various resolution levels for two different wavelet functions. The original trace is a *Ricker wavelet*, i.e. a normalized second-order derivative of a Gaussian function, with a peak frequency of 30 Hz. The left figure shows the decomposition by a Daubechies orthonormal basis with 2 vanishing moments, while the right figure shows the same decomposition with 3 vanishing moments. The Ricker wavelet (f) is the left most trace in each box, while the remaining traces correspond to f^j of equation (10), where $j = 1, 2, 3$, respectively. From Figure 2, it can be seen that the decomposed traces contains progressively lower frequencies with the increase of decomposition levels while the major features of the original signal are preserved. Comparing the two plots in Figure 2, we also observe that the increasing the number of vanishing moments increases the smoothness of the decomposed signal.

A Symmetric and Shift-Invariant Wavelet Basis

In many applications, it is required that the processes applied to the obtained signals be shift-invariant. For example, in examining the multi-scale property in residual statics correction problems, it is important that the error-fitting function

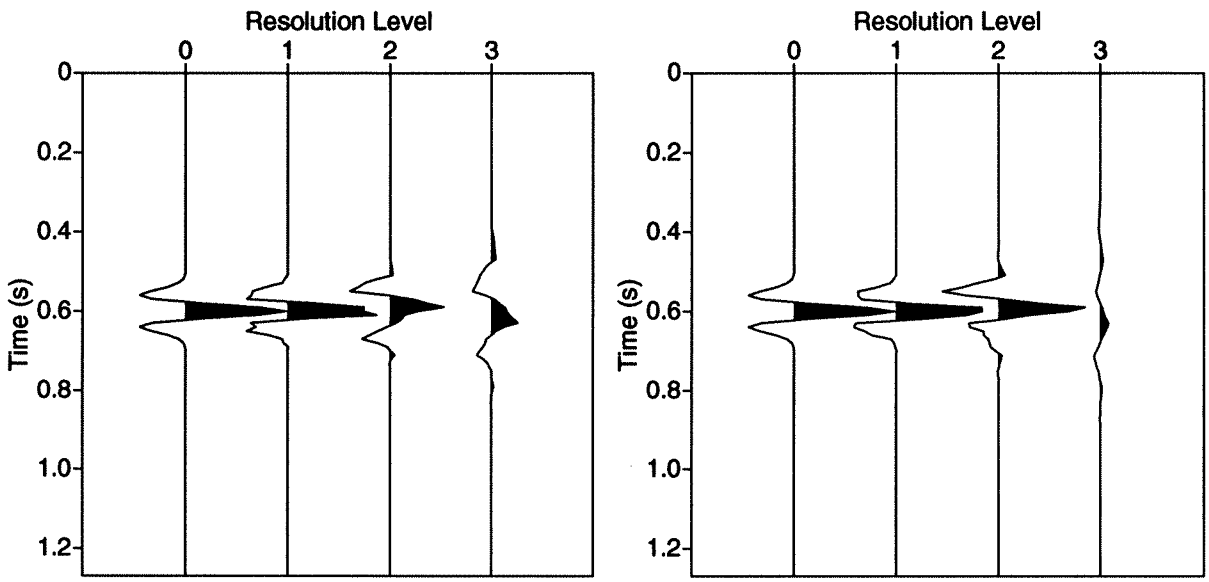

FIG. 2. Decomposition of a Ricker wavelet at increasingly coarser resolution levels. The bases of the decompositions are Daubechies wavelets with 2 and 3 vanishing moments for the left and right figure respectively. The first traces represents the signal at the finest level, which is the original signal.

at each scale have a common — or at least close to common — global minimum. Therefore, we expect that the relative time-shifts among traces at each scale to be almost the same as it was in the original data, and that the waveforms not be deformed from one trace to another. However, the orthonormal wavelet bases representations are generally not shift-invariant. This shift-variance can be seen directly from the construction of their bases, equations (2) and (4), because of the change of step sizes among different scales in these definitions. Therefore, the Daubechies wavelet bases are not suitable for our purpose. Figure 3 shows ten copies of randomly shifted Ricker-wavelet traces, and their projections onto the subspace V_3 in the Daubechies bases with 2 vanishing moments. The decomposed waveforms on the right of Figure 3 are deformed to different shapes among traces with different time-shifts, and they do not have the same relative time shifts of those shown on the left of Figure 3.

Saito and Beylkin (1993) suggested using the *shell* of an orthonormal basis when shift-invariant is required. Without loss of generality, let us assume that the signal we consider having finite length $N = 2^J$. Consider a family of functions

$$\{\tilde{\psi}_{j,k}(x)\}_{1\leq j\leq J,\,0\leq k\leq N-1}$$

and

$$\{\tilde{\phi}_{j,k}(x)\}_{1\leq j\leq J,\,0\leq k\leq N-1},$$

where

$$\tilde{\psi}_{j,k}(x) = 2^{-j/2}\psi(2^{-j}(x-k)), \tag{13}$$

$$\tilde{\phi}_{j,k}(x) = 2^{-j/2}\phi(2^{-j}(x-k)), \tag{14}$$

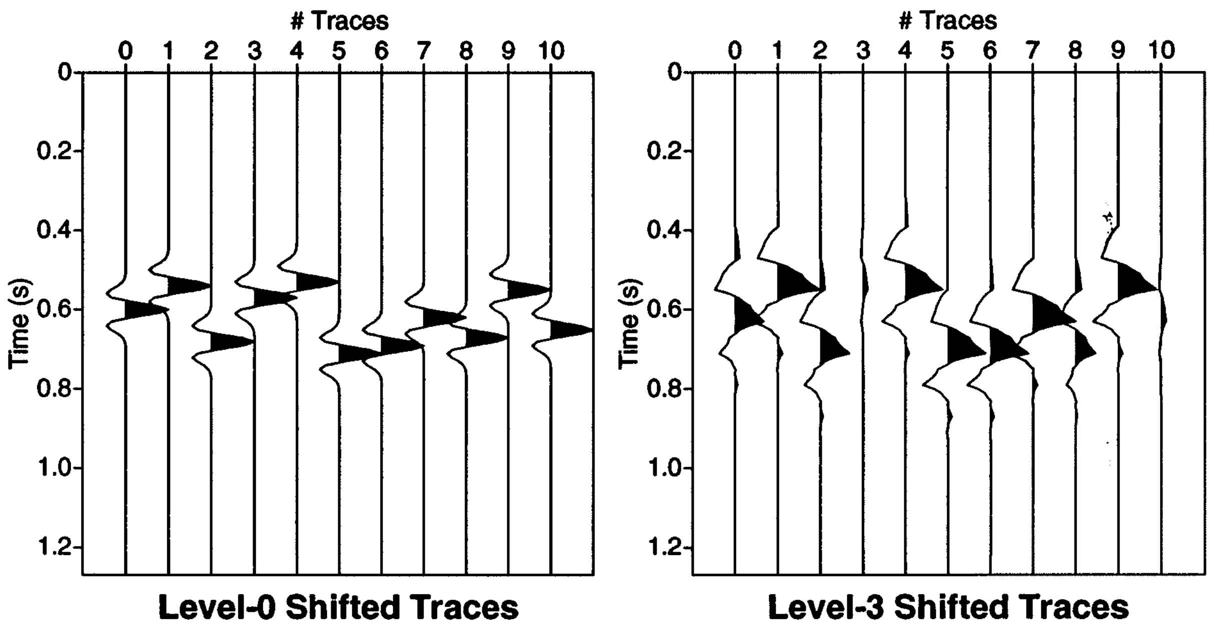

FIG. 3. Ten traces of randomly shifted Ricker-wavelet traces (left) and their decomposition at resolution level 3 in the Daubechies wavelet bases with vanishing moments of 2 (right).

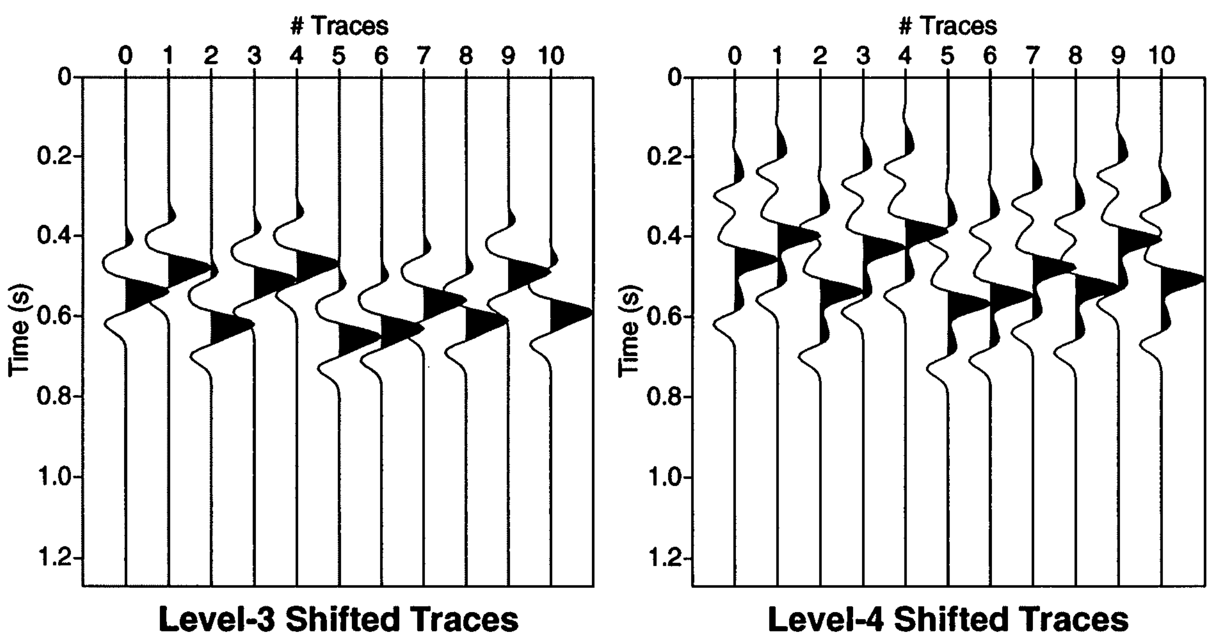

FIG. 4. The decomposition of ten copies of randomly shifted Ricker-wavelet traces, in the **shell** of the Daubechies basis with 2 vanishing moments, at resolution levels 3 and 4. The original traces are shown on the left of Figure 3.

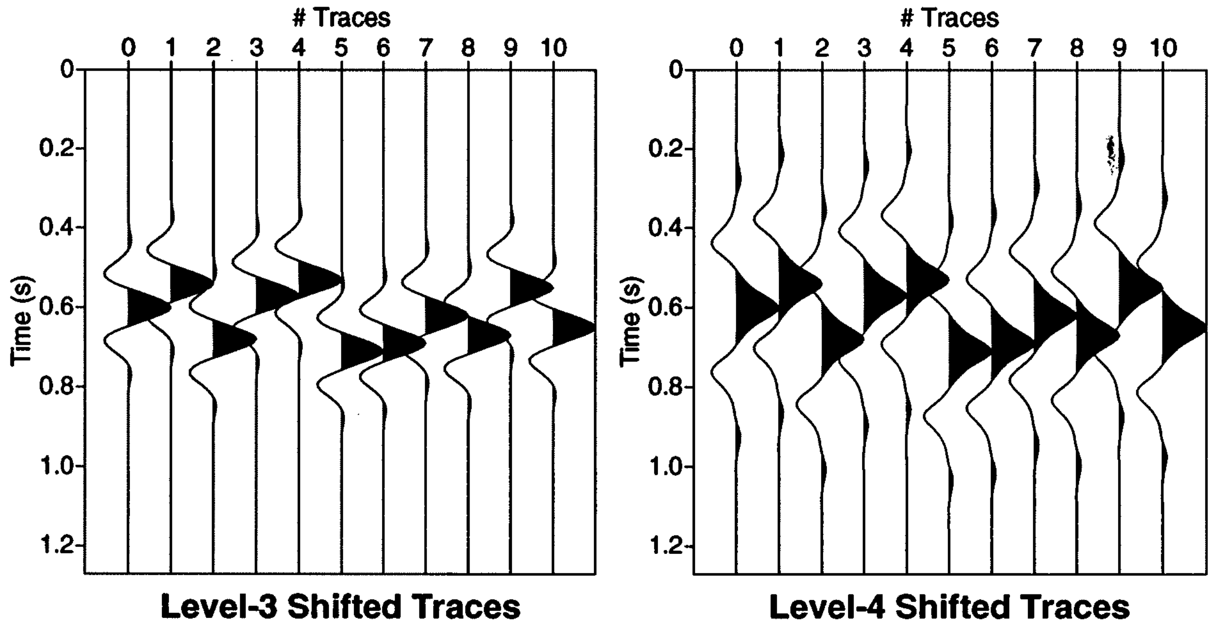

FIG. 5. The decomposition of ten copies of randomly shifted Ricker-wavelet traces, in the **auto-correlation shell** of Daubechies basis with 2 vanishing moments, at resolution levels 3 and 4. The original traces are shown on the left of Figure 3.

where the functions $\psi(x)$ and $\phi(x)$ are a wavelet and scaling function, respectively. The new family of functions defined by equations (13 and (14 can also serve as bases for subspaces V_j and W_j in MRA. They are complete, but they are redundant and not orthonormal (Saito, 1994). Therefore, the decomposition of a function in these bases is not unique. However, by forcing an additional constraint to the projection, a function $f \in V_0$ may still be decomposed in the shell of an orthonormal basis much the same way as it was in an orthonormal wavelet basis itself. In this case, the basis functions in equations (10) and (11) are replaced by $\tilde{\psi}_{j,k}(x)$ and $\tilde{\phi}_{j,k}(x)$.

The representation of signals using this family of bases are shift-invariant among different scales. Figure 4 shows the same numerical experiment as that in Figure 3, except using the shell of orthonormal bases expansion at resolution levels 3 and 4. The relative time-shifts among traces are preserved while the waveforms are deformed to the same amount. However, the original symmetric waveforms are deformed to asymmetric waveforms. This deformation of the waveforms is not desirable, and may cause problems for some applications.

To overcome this problem, a family of symmetric, shift-invariant bases are introduced (Saito and Beylkin, 1993). Let $\Phi(x)$ and $\Psi(x)$ be auto-correlation functions of scaling function and wavelet function respectively,

$$\Phi(x) = \int \phi(y)\ \phi(y - x)\ dy, \tag{15}$$

$$\Psi(x) = \int \psi(y)\ \psi(y - x)\ dy, \tag{16}$$

where ψ and ϕ satisfy equations (2) and (4) respectively. Construct a family of bases

$$\{\Psi_{j,k}(x)\}_{l\le j\le J,\,0\le k\le N-1}$$

and

$$\{\Phi_{j,k}(x)\}_{l\le j\le J,\,0\le k\le N-1},$$

where

$$\Phi_{j,k}(x) = 2^{-j/2}\Phi(2^{-j}(x-k)), \tag{17}$$

$$\Psi_{j,k}(x) = 2^{-j/2}\Psi(2^{-j}(x-k)). \tag{18}$$

Now, we have an *auto-correlation shell* of an orthonormal basis that is both symmetric and shift-invariant. Figure 5 shows the expansion of shifted Ricker-wavelet traces in the auto-correlation shell of Daubechies basis. It can be seen that both the symmetry of the waveforms and the relative time-shifts are preserved at resolution levels 3 and 4.

There exists a fast algorithm for expanding a function $f \in V_0$ using the auto-correlation shell of orthonormal basis (Saito and Beylkin, 1993). I only give the formulas of the discrete expansion; detailed derivation can be found in (Saito and Beylkin, 1993).

Suppose that S_k^j and D_k^j are the projected signal onto the subspaces V_j and W_j at the sampled positions respectively, that is

$$S_k^j = f^j(k\Delta), \quad D_k^j = Df^j(k\Delta),$$

where Δ is the sampling interval. Then, two symmetric filters, $P = \{p_k\}_{-L+1\le k\le L-1}$ and $Q = \{q_k\}_{-L+1\le k\le L-1}$ are applied recursively to the signal we wish to decompose,

$$S_k^j = \sum_{l=-L+1}^{L-1} p_l\, S_{k+2^{j-1}l}^{j-1}$$

$$D_k^j = \sum_{l=-L+1}^{L-1} q_l\, S_{k+2^{j-1}l}^{j-1}; \tag{19}$$

$$\tag{20}$$

where $0 \le k < N$, $1 \le j \le J$, and L is the filter length in the "two-scale difference" equations of wavelet and scaling functions as in equations (2) and (4). In equation (20), $N = 2^J$ is the number of samples of the signal and the filter coefficients p_k and q_k are,

$$p_k = \begin{cases} 2^{-1/2}, & \text{for } k = 0, \\ 2^{-3/2}a_{|k|}, & \text{otherwise;} \end{cases} \tag{21}$$

and

$$q_k = \begin{cases} 2^{-1/2}, & \text{for } k = 0, \\ -p_k, & \text{otherwise.} \end{cases} \tag{22}$$

In equations (21) and (22), coefficients $\{a_k\}_{k=1,\ldots,L-1}$ are the correlation of the low-pass filter $\{h_l\}_{l=0,\ldots,L-1}$ in equation (4),

$$a_k = \begin{cases} 2\sum_{l=0}^{L-1-k} h_l\, h_{l+k}, & \text{for } k \text{ is odd}, \\ 0, & \text{for } k \text{ is even}. \end{cases} \tag{23}$$

MULTIRESOLUTION ANALYSIS FOR INVERSE CALCULATIONS

Many inverse problems are solved by optimization methods. Mathematically, gradient-search optimization methods work well when the objective function is convex (e.g., a "basin") in the searching range; and the wider the basin of attraction leading to the bottom, the more likely that the optimizations converge to the optimum point. Optimizations have difficulties when there exists more than one point with zero gradient (e.g. local minima, flat area) in the searching range. Unfortunately, this is usually the case in many realistic inverse problems. The complexity of objective functions can be affected by many factors, such as noise, frequency bandwidth, and features of the information in the observed data.

As studied in the above section, an MRA can decompose signals into various resolution levels. The data with coarse resolutions contain less detailed information and lower frequencies, while keeping major features of the original signal consistent with the low frequency information. These less-information data can serve as a relaxation to optimizations. Therefore, by using data at coarser resolution levels, complexity of objective functions may be reduced, which increases the performance of optimizations.

A Simple Residual Statics Problem

Let us first consider a simple residual statics problem. Consider a trace containing one Ricker wavelet; duplicate the trace with an unknown shift. Figure 6 shows two traces as described above. Now, we look for the time-shift between the two traces by applying an optimization, that is, searching for the time-shift which maximally aligns the two traces. This is a simple residual statics estimation problem using the stacking power method; there is only one unknown in the optimization. The objective function is formulated as a least-squared error,

$$E(\delta) = \sum_{i=0}^{N-1} (P_0(i-\delta) - P_1(i))^2, \tag{24}$$

where $P_0(t)$ and $P_1(t)$ are the two data traces, N is the number of samples per trace, and δ is the unknown time-shift. The goal is to find the time-shift δ that minimizes the error function $E(\delta)$. Figure 7 shows the error function as in equation (24) for the fitting of these two traces. In addition to possible problems caused by the local-minima, the basin of attraction leading to the global-minimum is "steep" and narrow, while the two areas to the sides are "flat". The global structure of this objective function suggests that the global minimum point may be hard to find by traditional

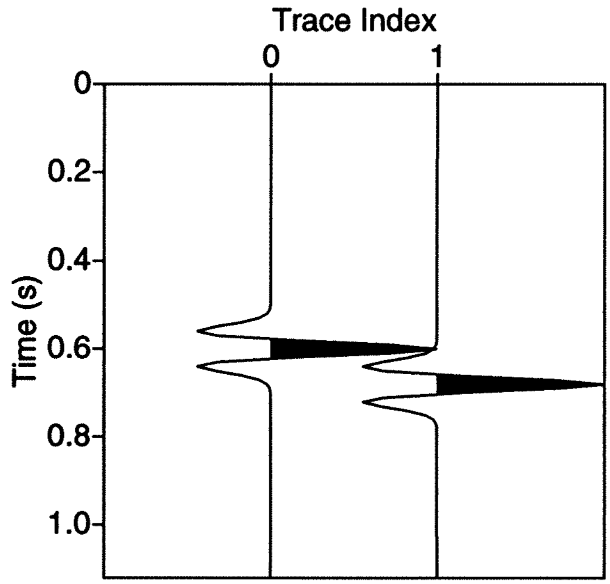

FIG. 6. The observed data in the first example. Two traces contains identical waveforms of Ricker wavelet with a 30 Hz peak frequency. The relative time-shift is the unknown we are seeking.

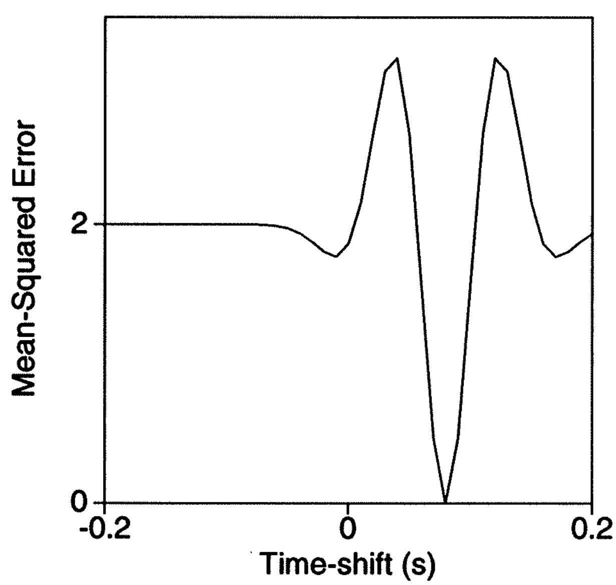

FIG. 7. The error-fitting function with respect to the relative time-shift between two traces. The goal is to find the optimal point where the mean-squared error is minimum.

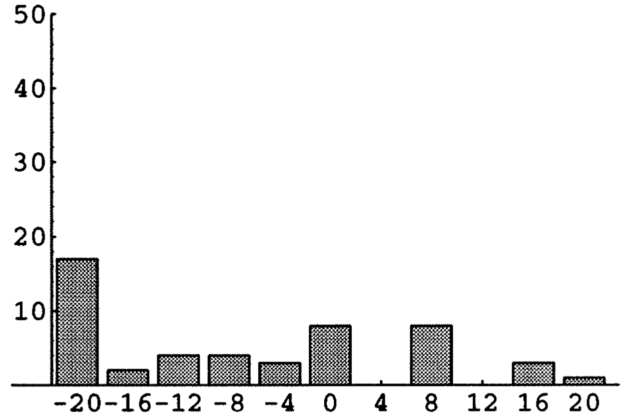

FIG. 8. The histogram of the obtained time-shifts of 50 conjugate-gradient optimization experiments starting from uniformly distributed random initial models between $[-0.2, 0.2]$ s. The horizontal axis is the number of shift-samples, where the sample interval is 0.01 s, and the grid size of the histogram is 4 samples. The number of times that found the true global minimum is 8 out of 50.

gradient-based searching methods. Assuming that we know *a priori* the time-shift between the traces lies in the range of $[-0.2, 0.2]$ s, the searching range is restricted to this interval. Figure 8 shows the histogram of the obtained time-shift for 50 optimizations by using the **Conjugate-Gradient** and **Cubic-Line-Search** tools provided in the CWP Object-Oriented Optimization Library (Deng et al., 1995); initial models are randomly chosen between $[-0.2, 0.2]$ s. As expected, the chances of finding the correct global minimum is small. In the case of this test, there are 8 out of 50 experiments that the correct time-shift was found.

Using the MRA for Optimizations

Let us decompose the observed data into various resolution levels by representing them with wavelet bases. For the above example, the traces $\{P_i(x);\ i = 0, 1\}$ of lengths $N = 2^J$ can be represented in the form of equation (12),

$$P_i(x) = P_i^{n_0}(x) + \sum_{j=n_0}^{1} \sum_{k=0}^{N-1} d_{j,k}\ \psi_{j,k}(x), \tag{25}$$

where $1 \leq n_0 \leq J$ and $P_i^{n_0}(x)$ is the projection of the original data onto the subspace V_{n_0}. Therefore, equation (24) can be rewritten as,

$$E^{n_0}(\delta) = \sum_{i=0}^{N-1} (P_0^{n_0}(i - \delta) - P_1^{n_0}(i))^2 + R^{n_0}(\delta), \tag{26}$$

where $R^{n_0}(\delta)$ is the residual error term which is related to the detailed information being projected onto subspaces $\{W_j;\ j = 1, ..., n_0\}$.

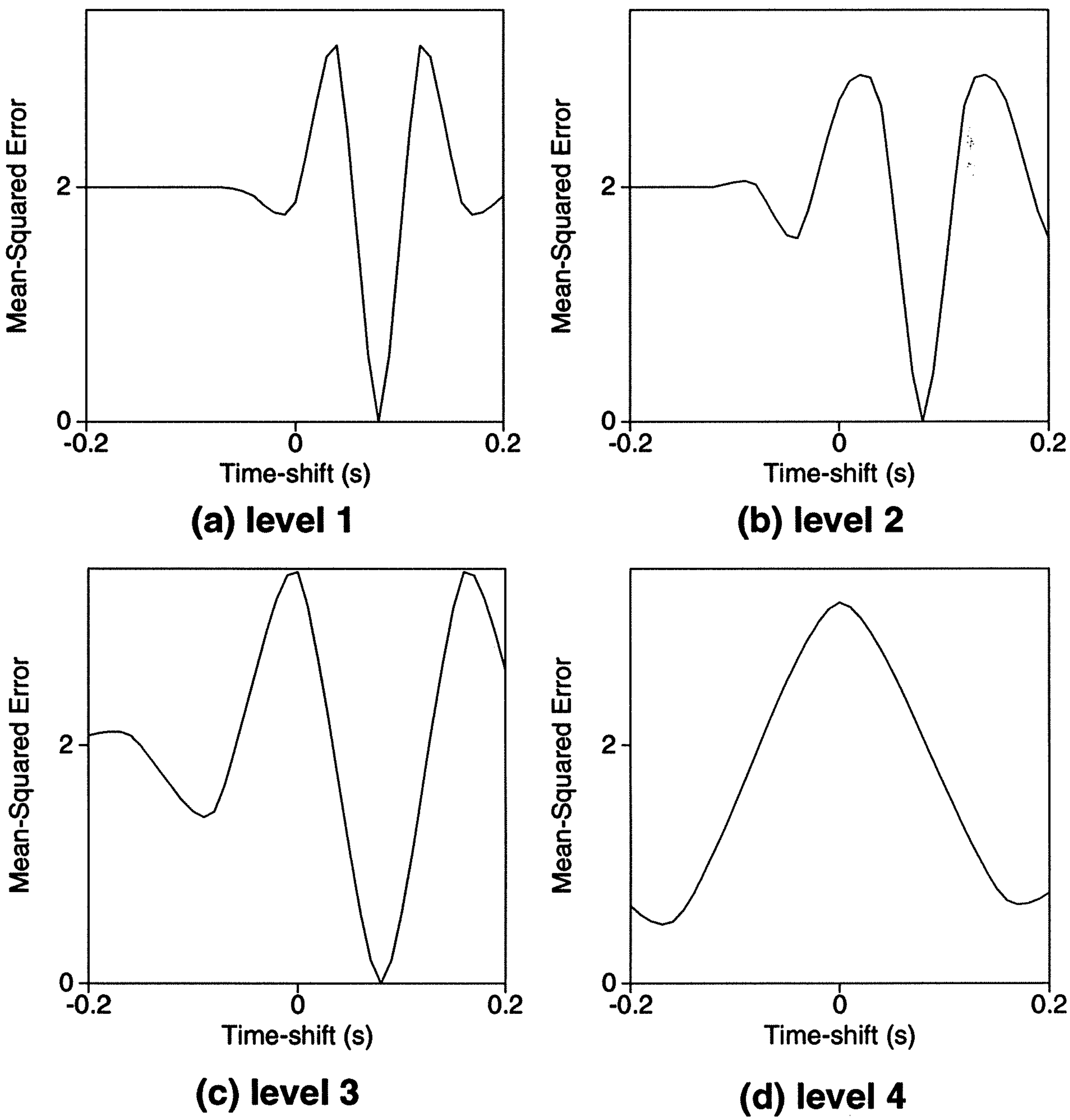

FIG. 9. The mean-squared error functions for two seismic traces at various resolution levels. The traces are decomposed in the Daubechies basis with 2 vanishing moments.

Ignoring certain levels of fine-resolution information, i.e., ignoring the residual term in equation (26), the resolution level n_0 representation of the seismic traces can be used for optimization. Figure 9 shows the objective function $E^{n_0}(\delta)$ at various resolution levels, $n_0 = 1, 2, 3, 4$. It shows that the global complexity of the objective function is reduced with the increasingly coarser level of resolution, and there are wider basins of attraction leading to the global minimum. However, Figure 9(d) shows that the global structure of the objective function is severely distorted when detailed information is ignored.

This phenomenon is caused by the shift-variance nature of compactly supported, orthonormal wavelet bases. In the inverse problem discussed here, it is required that the bases used to represent the signals be shift-invariant. According to the discussion of the previous section, two families of bases are shift-invariant. Because of the symmetric feature of the auto-correlation shell of orthonormal bases, we choose this family of bases for this study. From this point on, the paper uses only the auto-correlation shell of orthonormal bases to decompose the signal, unless otherwise indicated. Figure 10 shows the objective function at resolution levels $n_0 = 2, 3, 4, 5$ in an auto-correlation shell of the Daubechies wavelet basis with 2 vanishing moments. The global structure of the objective function also shows the desired simplification as that in Figure 9, such as a wider basin of attraction leading to the global minimum, less oscillations and smaller "flat" area in the searching range. Moreover, the global minimum is not shifted at any decomposed resolution level. Figure 10(d) show that the whole searching range transformed to one wide basin of attraction, which would lead all initial models to the global minimum.

Figure 11 shows the same histograms as that shown in Figure 8, except the data used for optimizations are decomposed at various resolution levels. These results confirm our prediction that there are increasing chances for local-search optimizations to find the global minimum when coarse-resolution data are used. For Figure 11(d), all searches converge to the global minimum when data are decomposed to resolution level five.

For the simple problem discussed above, five levels of decomposition are needed to reduce the objective function to a convex function in the searching range. In addition, the global minimum of this simplified objective function coincides with that of the original objective function. Therefore, the correct solution is reached when only coarse resolution data are used in this example. In the next example, an optimization applied to the coarse-resolution data will not suffice.

More Examples

For more complex optimization problems, further reduction of resolution may be needed to make objective functions convex. The severe loss of information may cause an erroneous global minimum of the objective function.

Here, I show another example of residual statics correction problem for a trace with complex waveforms and unknown noise. Figure 12 shows a trace taken from a

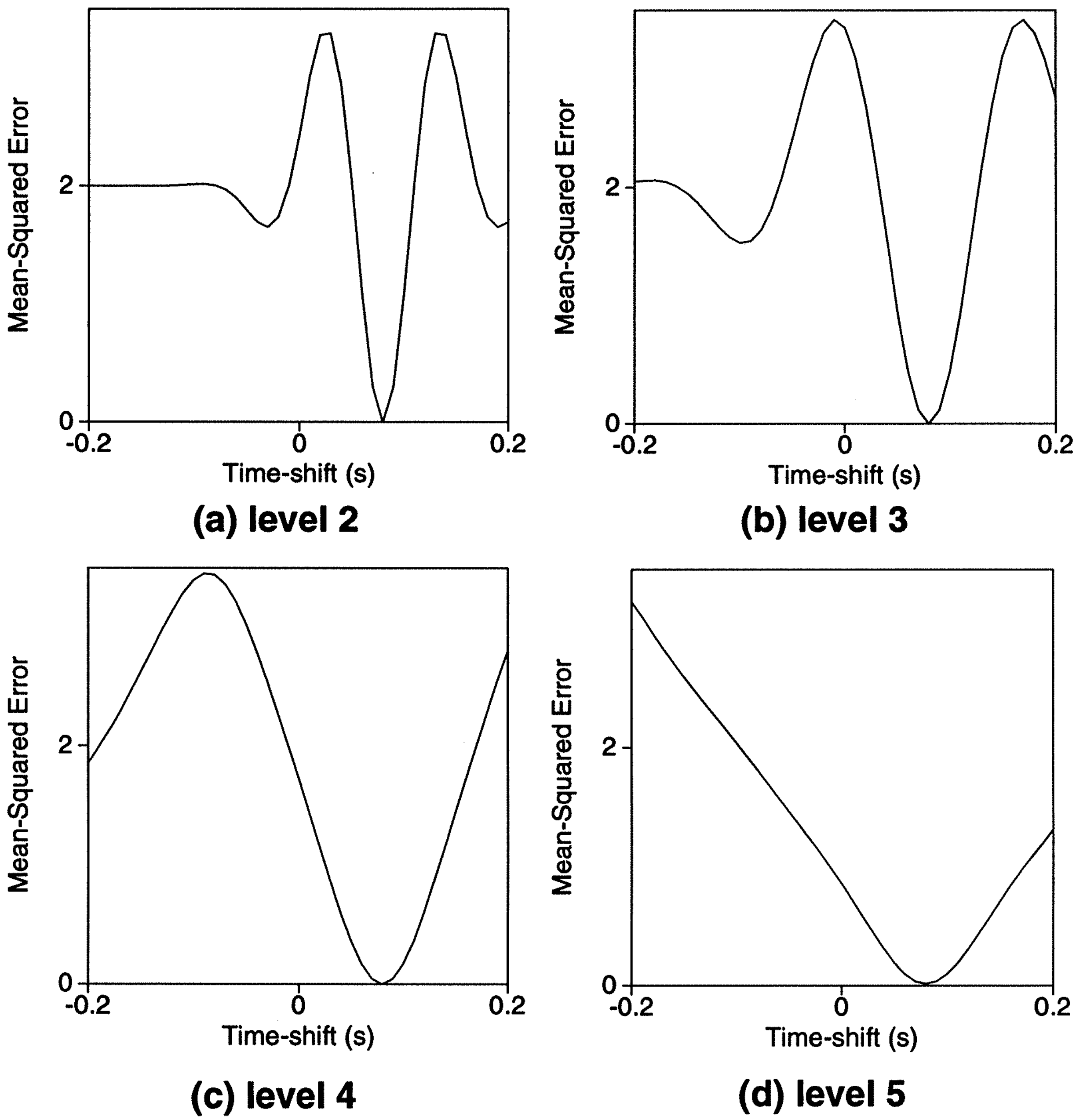

FIG. 10. The mean-squared error functions for two seismic traces at various resolution levels. The traces are decomposed in the auto-correlation shell of the Daubechies basis with 2 vanishing moments.

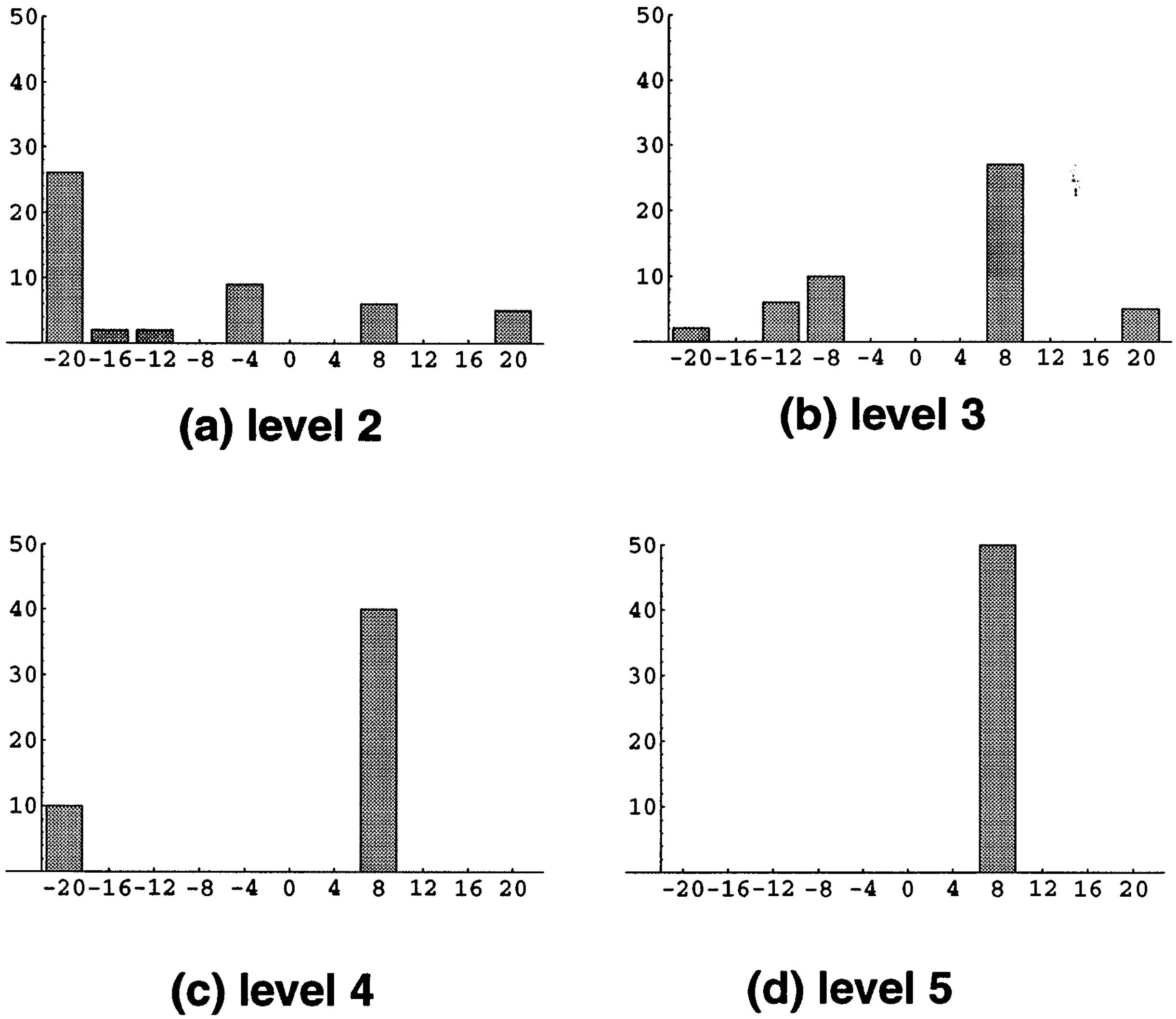

FIG. 11. Histograms of the obtained time-shifts of 50 conjugate-gradient optimization experiments for data at various resolution levels. Initial models are chosen randomly between $[-0.2, 0.2]$ s. The horizontal axis is the number of shift-samples, where the sample interval is 0.01 s, and the grid size of the histograms is 4 samples. All 50 experiments found the true solution when the data are decomposed to resolution level 5.

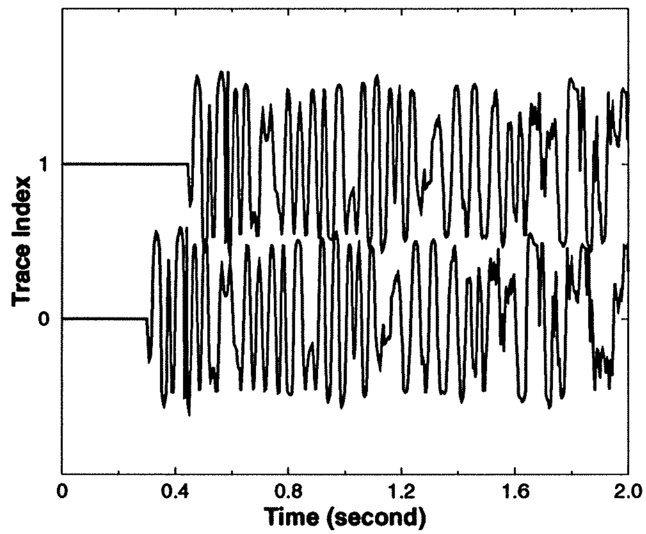

FIG. 12. A real seismic trace and its duplication with an unknown shift.

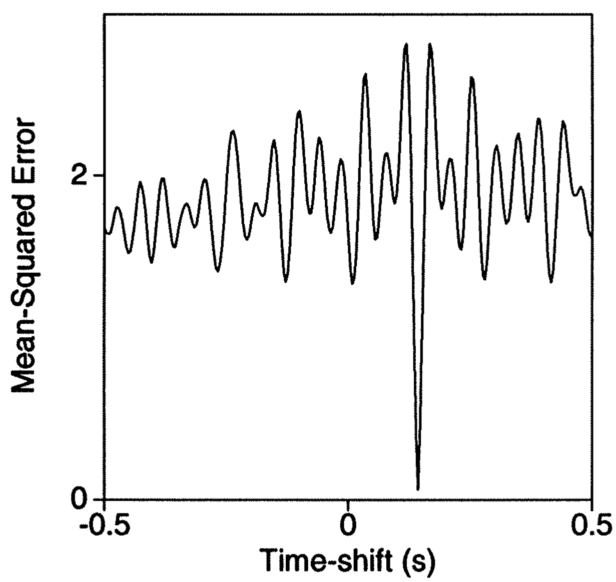

FIG. 13. The mean-squared error functions for two real seismic traces shown in Figure 12.

field seismic record, and its duplication with an unknown shift. We repeat the process discussed in the previous section on these two traces. Figure 13 shows the objective function for this optimization. Due the oscillatory nature of the seismic field data and unknown noise, the objective function shows complicated local and global structure. The basin of attraction leading to the global minimum point is extremely narrow and steep, which makes it almost impossible for any gradient searching methods to find the correct solution.

Again, the auto-correlation shell of the Daubechies basis is used to decompose the traces to coarse resolution levels. Figure 14 shows objective functions when applying various level of decomposition to traces in Figure 12. As expected, the complexity of the objective function is greatly reduced after the data being decomposed to coarse levels.

However, it is worth noticing the global minimum point are slightly shifted in Figure 14(d), though the objective function shows a nice, convexity shape. This problem may be caused by the loss of information when too much resolution was discarded from the data. In this case, an iterative process similar to a multi-grid iteration can be used to enhance the resolution progressively; i.e. the solution of a coarse-level optimization is used as the initial model to the following optimization at a finer level (e.g., Chen 1994).

DISCUSSION

The MRA can be used for analyzing signals at various scales. One of these first studies in the field of wave propagation was conducted by Morlet et al. (1982a; 1982b). In recent years, many researchers have been applying the technique successfully to data compression and processing. Cohen and Chen (1993) gave some intuitive insight as well as suggestions on possible applications in seismic imaging. This paper shows that wavelet theory can also be used to study optimization as applied to inverse theory.

Taking advantage of MRA in wavelet theory, seismic data can be decomposed to coarse resolution levels, while keeping major features in the original signal. This paper has described the first step of the study on the complexity of inverse calculations; the influence of MRA on objective functions has significant effects on the performance of optimizations. The objective functions can be simplified and they approach convexity when data are decomposed to a low resolution. This initial study demonstrates that MRA may be a useful tool for characterizing complexity of objective functions in the optimization approach to inverse problems.

Using the residual statics correction as an example, I have shown in this report that the waveform misfit function is multi-modal even for a simple problem, and MRA indeed can simplify the complexity waveform misfit function. Comparing Figure 7 and Figure 10, Figures 13 and 14, the objective functions for the wavelet decomposed data show a wider basin of attraction leading to the global minimum, a reduced

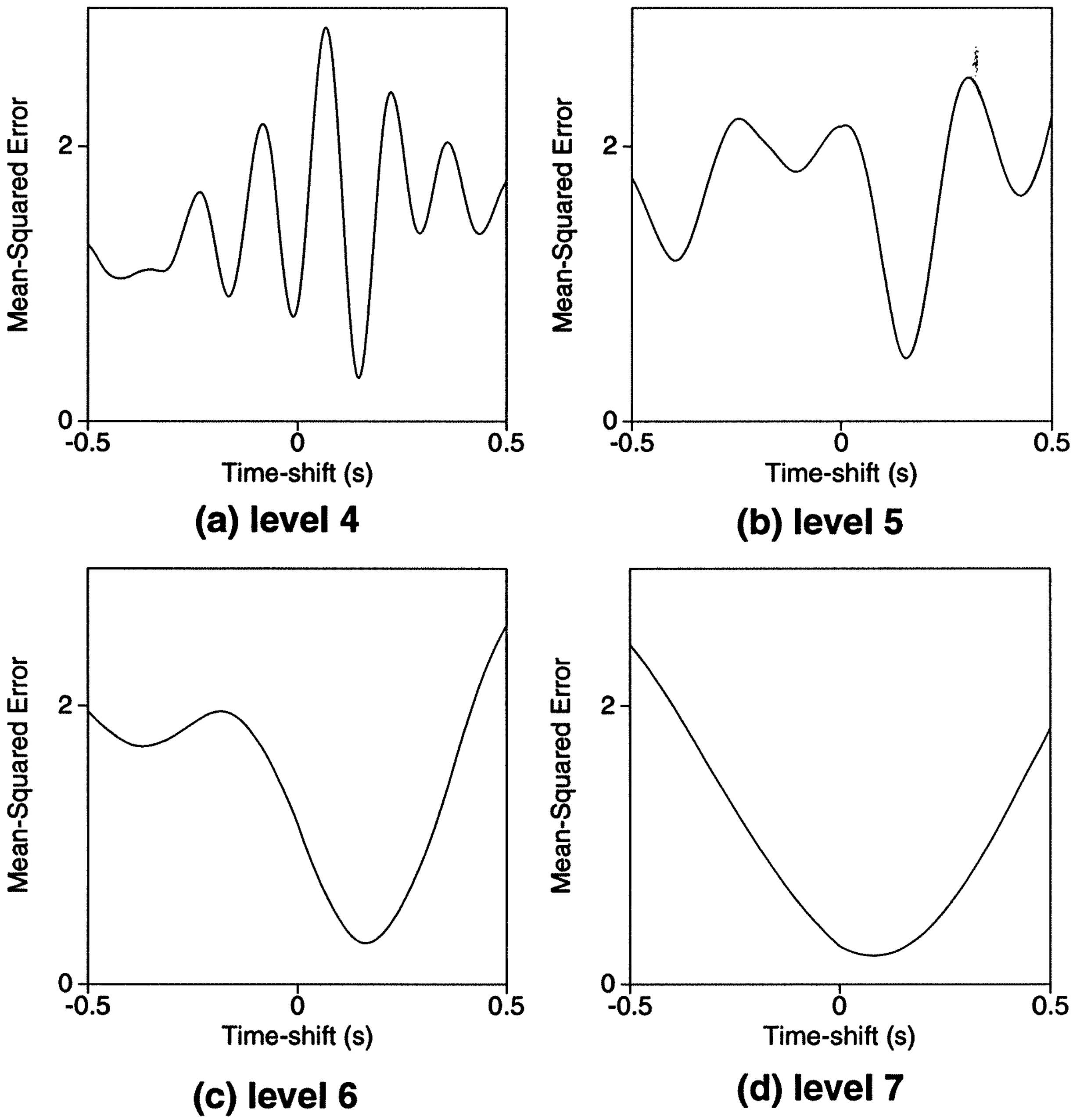

FIG. 14. The mean-squared error functions for two real seismic traces shown in Figure 12 at various resolution levels. The traces are decomposed with the auto-correlation shell of the Daubechies basis with 2 vanishing moments.

number of local minima, but non-distorted global feature.

As discussed in this paper, the choice of bases used to decompose the data is critical for this application. In addition to the time-invariance and the symmetry issues of the bases, the influence of vanishing moments to objective functions remains to be investigated, especially its tradeoff with computational intensity.

More tests will be done for more realistic optimization problems. For example, inverse problems we encounter usually have many unknown parameters to be recovered; the observed data may also be contaminated by noise. Therefore, it is important to study the influence of the MRA on these realistic problems. Results of these studies can be used to characterize complexities of certain inverse problems. As all the other inverse algorithms, the computation cost is also an issue that needs to be studied.

ACKNOWLEDGMENT

I appreciate very much the encouragement of Dr. John Scales during the progress of this project, his continuous support on the optimization library – COOOL, and sharing his insights in inversions. I also would like to thank Drs. Bin Wang and Keh Pann of Mobil MEPTEC for initially suggesting that I look at the wavelet transform literature. I benefited a great deal from the "wavelets" class I took from Dr. Gregory Beylkin this spring. I am sincerely grateful to him for offering this wonderful opportunity and his suggestion of looking at the **symmetric, shift-invariant** wavelet bases. I also benefited from the brief discussion with Dr. Stew Levin of Mobil MEPTEC during SEG, Los Angeles.

Finally, I want to thank my colleague and friend, Wences Gouveia, who helped with the optimization algorithm part of this project.

REFERENCES

Cary, P., and Chapman, C., 1988, Automatic 1-D waveform inversion of marine seismic refraction data: Geophysical Journal, **93**, 527–546.

Chapman, C., and Orcutt, J., 1985, Least squares fitting of marine seismic refraction data: Geophysical Journal of the Royal Astronomical Society, **82**, 339–374.

Chen, T., 1994, Multilevel differential semblance optimization for waveform inversion: Technical Report CWP-153, Center for Wave Phenomena, Colorado School of Mines.

Chevent, G., 1994, Duality methods for waveform inversion: Preprint.

Cohen, J. K., and Chen, T., 1993, Fundamentals of the wavelet transform for seismic data processing: Technical Report CWP-130, Center for Wave Phenomena, Colorado School of Mines.

Daubechies, I., 1992, Ten lectures on wavelets: SIAM.

Deng, H., Gouveia, W., and Scales, J. A., 1995, The CWP Object-Oriented Optimization Library: submitted to The Leading Edge.

Gouveia, W., 1993, A study on distributed genetic algorithms: Technical Report CWP-137, Center for Wave Phenomena, Colorado School of Mines.

Gouveia, W., 1994, Residual statics estimation by a distributed genetic algorithm: 64th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Soc. Expl. Geophys., 64th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 195–199.

Jawerth, B., and Sweldens, W., 1994, An overview of wavelet based multiresolution analyses: SIAM Review, **36**, 377–412.

Mallat, S., Jul. 1989, A theory for multiresolution signal decomposition: IEEE Trans. Pattern Anal. Machine Intell., **11**, 674–693.

Meyer, Y., 1992, Wavelets and Operators: Cambridge University, Cambridge.

Monzon, L. A., May 1994, Constructive multiresolution analysis and the structure of quadrature mirror filters: Ph.D. thesis, Yale University.

Morlet, J., Arens, G., Fourgeau, E., and Giard, D., 1982a, Wave propagation and sampling theory – Part I: Complex signal and scattering in multilayered media: Geophysics, **47**, 203–221.

——— 1982b, Wave propagation and sampling theory – Part II: Sampling theory and complex waves: Geophysics, **47**, 222–236.

Rothman, D., 1985, Nonlinear inversion, statistical mechanics and residual statics estimation: Geophysics, **50**, 2784–2796.

Rothman, D., 1986, Automatic estimation of large residual statics corrections: Geophysics, **51**, 332–346.

Saito, N., and Beylkin, G., 1993, Multiresolution representations using the autocorrelation functions of compactly supported wavelets: IEEE Transactions on Signal Processing, **41**, 3585–3590.

Saito, N., December 1994, Local feature extraction and it's applications using a library of bases: Ph.D. thesis, Yale University.

Saleck, F. M., Bunks, C., Zaleski, S., and Chavent, G., 1993, Combining the multigrid and gradient methods to solve the seismic inversion problem: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Soc. Expl. Geophys., 688–691.

Scales, J. A., and Tarantola, A., 1994, An example of geologic prior information in a bayesian seismic inverse calculation: Technical Report CWP-159, Center for Wave Phenomena, Colorado School of Mines.

Scales, J. A., Smith, M. L., and Fischer, T. L., 1992, Global optimization methods for multimodal inverse problems.: Journal of Computational Physics, **103**, 258–268.

Sen, M., and Stoffa, P., 1991a, Nonlinear one dimensional seismic waveform inversion using simulated annealing: Geophysics, **56**, 1624–1638.

——— 1991b, Rapid sampling of model space using genetic algorithms simulated annealing: Geophysical Journal International, **56**, 27–29.

Sen, M., and Stoffa, P., 1992, Genetic inversion of AVO: The Leading Edge, **11**, 27–29.

Shaw, P., and Orcutt, J., 1985, Waveform inversion of seismic refraction data and applications to young pacific crust: Geophysical Journal of the Royal Astronomical Society, **82**, 374–414.

Smith, M., Scales, J., and Fischer, T., 1992, Global search and genetic algorithms: The Leading Edge of Exploration, **11**, 22–26.

Symes, W., and Carazzone, J., 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Symes, W., 1993, A differential semblance criterion for inversion of multioffset seismic reflection data: J. Geophysical Research B, **98**, 2061–1074.

Symes, W., 1994, Convex duality and separable inverse problems in wave propagation: Technical Report TR92-22, Department of Computational and Applied Mathematics, Rice University.

Tarantola, A., 1987, Inverse problem theory: Elsevier, New York.