Noise Attenuation: A Hybrid Approach Based on Wavelet Transform

Rongfeng Zhang and Daniel Trad - University of British Columbia

CSEG Geophysics 2002

Abstract:
Field seismic data can contain many sorts of noise and interferences. If the noise is stronger in order of magnitude than the signal, most techniques often applied in seismic processing will be severely affected. Denoising methods, even very robust schemes such as physical wavelet frame denoising (PWFD), are not exceptions. In this paper, we present a robust, data adaptive and fast 1D wavelet transform (WT) method to attenuate this kind of noise and combine it with the 2D PWFD method to achieve a hybrid strategy for noise attenuation.

Introduction

Strong noise in field seismic data sometimes can be several orders of magnitude larger than the signal. Attenuating this kind of noise is very important for the afterwards processing. The wavelet transform (WT) has been largely used to attenuate noise in image and signal processing, including geophysics. In most cases it is used to depress random noise, following a thresholding theory developed by Donoho (1993). The idea is that, given a signal with random noise, this noise will map mainly to small WT coefficients because its energy is distributed along all scales. The signal, on the contrary, maps to large WT coefficients because it maps to some scales more than others, due to its coherence. Therefore a simple scheme to attenuate noise is to compute adequate thresholds for every scale and zero out or downweight the coefficients below these thresholds. This procedure has been successfully applied in many fields and several statistical estimators for the thresholds have been formulated for the case of Gaussian noise. However, if the noise is not Gaussian, coherent and especially of strong amplitude, the above criteria will break. In this situation, we must be careful to explore denoising methods to avoid signal distortion.

In this paper, we introduce a hybrid two-step approach to attenuate high amplitude noise in seismic gathers. First, we calculate the 1D WT of the data and, instead of removing low coefficients, we remove or attenuate very large WT coefficients. This filtering process is performed scale by scale, modifying the threshold values at every scale. This kind of filtering has been originally applied to filter impulsive noise in magnetotelluric data [Trad and Travassos2000]. When applying this procedure to seismic data, strong coherent noise is, in general, not totally removed, but partly attenuated leaving signal untouched. By applying in a second step the 2D wavelet frame denoising filtering [Zhang2000] both coherent and random noise are depressed.

Discrete WT filtering

Wavelets are used to represent a time series in the same way as trigonometric functions in Fourier Analysis. One important difference is that in wavelet analysis the scale in which we look at the data plays a crucial role: wavelets process data at different scales, or resolutions.

In wavelet analysis one adopts a wavelet prototype function called the analyzing or mother wavelet. This work uses the Daubechies wavelet [Daubechies1992] as the analyzing wavelet. This is an orthogonal, fractal wavelet with a compact representation. The class of orthogonal wavelets is widely used in multiresolution analysis. As any particular set of wavelets, the Daubechies is specified by a set of coefficients. In particular we have generated this set of wavelets with 4 to 20 coefficients. The best results in this work were obtained with the Daubechies wavelet with 20 coefficients, which is a very smooth wavelet. This is so because the strong interferences in seismic gathers are often surface waves which have smooth waveform and low frequency.

Let $\psi(t)$ be the mother wavelet defined in the space of square-integrable functions over the real numbers $L^2(R)$. Through dilations and translations of $\psi(t)$ one constructs an orthogonal basis

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \tag{1}$$

where $a \neq 0$, $b \in R$. The scale factor a gives dilations and compressions and b is a translation in time. Varying a gives rise to a spectrum. The translation factor b represents the sliding of $\psi(t)$ over a given time series F(t). The continuous wavelet transform of a time series $F(t) \in L^2(R)$ is defined as

$$f(a,b) = \langle \psi_{a,b}(t), F(t) \rangle = \int_{-\infty}^{\infty} \psi_{a,b}(t)\, F(t)\, dt, \tag{2}$$

The time series F(t) can be fully recovered from the transform (2) by the reconstruction relation, also called resolution of the identity [Vetterli and Kovacevic1995]. That relation states that any time series F(t) can be written as a superposition of shifted and dilated wavelets.

The seismic traces are transformed into the wavelet-domain through a fast Discrete Wavelet Transform (DWT) algorithm [Press 1992]. Assume a time series F(t) of a seismic trace, with $N=2^{J+1}$ data points. For the sake of simplicity n_s will be used to denote the number of coefficients at the resolution level a. The resolution level , namely level $a=J$, has $n_s=2^J$ wavelet coefficients. In the remainder of this section we assumed that all the data have already been transformed into the wavelet-domain.

Begin by estimating a scale for the wavelet coefficients. This scale will be needed for thresholding in the filtering process. Assume a zero mean Gaussian noise with a small fraction of outliers. If most of the wavelet coefficients $f(a,b)$ represent useful signal, then the median absolute deviation (MAD) can be used as a scale factor for the wavelet coefficients because it reflects the size of the original data [Rousseeuw and Leroy1987]. The MAD is a robust estimate for the standard deviation. It can be calculated as

$$s(a) = \frac{median\left\{\left|f_a(b) - \tilde{f}_a\right|\right\}}{\sigma}, \tag{3}$$

where $f_a(b)$ is the b -shifted wavelet coefficient at a fixed scale a, and $\tilde{f}_a = median\{f_a(b)\}$, for all possible shifts b. It is clear from relation (3) that the MAD depends on the resolution level a. The symbol σ stands for the theoretical MAD for the considered standard probability distribution. The value for σ considered in this paper is σ=0.6745, for a Gaussian distribution.

The next step is to design a threshold level in order to identify anomalous data. The thresholding scheme is based on the assumption that noise in the wavelet transform is also Gaussian and approximately stationary at each resolution level a. This follows from the fact that the wavelet basis is orthogonal. In the wavelet literature a threshold is usually set in order to discard values below a certain level. This work follows an inverse route as its objective is to eliminate energetic regional and strong interferences like air waves. Assuming that the useful part of the seismic signal behaves like a Gaussian distribution 1 in the time domain then so it remains in the wavelet-domain and with the same relative amplitude to isolated events. The energy of strong interferences is compressed into a very small number of large amplitude wavelet coefficients. Those coefficients stick up well above normal signal amplitude levels where the useful signal smears at all scales due to its white noise characteristics. We then set a threshold to identify and eventually weight down coefficients that are above it.

The thresholds will depend on the resolution level a in the same way as the scale does. Remember that n_s is the number of wavelet coefficients at a particular resolution level a. With the scale $s(a)$, relation (3), it is possible to construct a threshold for the wavelet coefficients. We consider a threshold for a Gaussian distribution, but other distributions can be considered. The threshold for a Gaussian distribution is [Donoho1993]

$$\tau = C_a \; s(a)\sqrt{2 \log n_a}\,, \tag{4}$$

where C_a is an empirical factor to account for the need to increase the threshold τ at lower resolution levels. At those levels the low frequency component of the data is represented with fewer coefficients. Therefore throwing out useful signal is more likely to occur than at the higher resolution levels where there are more coefficients. Note that only a few periods of signal may occur in the entire time series at the highest levels of decimation.

In this work we found that $C_a = \log N / \log n_s$ gives good results. Assuming that the useful signal looks like white noise its DWT coefficients will have a variance that grows roughly as 2a. The factor Ca prevents the heavy dumping of the square root in the thresholds τ, relation (4).

Any coefficient that sticks above the chosen threshold τ, relation (4), can be weighted down in order to produce the coefficients $f(a,b)$ that will be transformed back to the time domain. The choice here is to use a severe data-adaptive robust filter that falls off quickly or a softer down-weighting that shrinks the coefficients to the size of the thresholds. These two possibilities can be experimented with Thomson and Huber weights [Trad and Travassos2000]. The Thomson weight function is given by

$$\varpi_a(b) = \exp\left\{-\exp\left[\tau\left|\frac{f_a(b)}{s(a)}\right| - \tau\right]\right\}, \tag{6}$$

Relation (6) is smooth and falls off very quickly as soon as coefficients become larger than the thresholds in relation (4). The filtering algorithm computes weights from relation (6) for each scale, or resolution level a independently.

Huber weights are defined as

$$\varpi_a(b) = \begin{cases} \frac{s(a)}{f_a(b)}, & \text{if } f_a(b) > \tau \\ 1, & \text{if } f_a(b) < \tau. \end{cases} \tag{7}$$

The effect of these weights is just to shrink the coefficients, which are above the chosen threshold to τ. This second weight function has been employed in the examples of this paper with better results than Thomson weights.

PWFD

Wavelet frames are a powerful tool for the analysis of data. In as much as a frame is composed of atoms or wavelets, the observed data can also be regarded as being composed of atoms or wavelets. It is intuitively clear that, the more similar are the analysis and data atoms to each other, the more efficient in terms of parsimony is the method of signal representation, and the more useful information can be retrieved from the data.

Seismic data, fortunately for us, exhibit high trace to trace coherence. Denoising methods, including wavelet applications, that are based on individual traces, cannot, therefore, attenuate the noise that is present, in an optimal manner. Conventional 2D wavelet denoising methods use a 2D wavelet [Miao and Cheadle1998,Nguyen and Mars1999] and therefore attempt to take into account the relationship that exists between traces. Since these implementations typically use a tensor multiplication of 1-D wavelets, the emphasis is on the vertical, horizontal and diagonal dependence in the data. If an angle parameter is used, directions corresponding to this angle can also be taken into account. In this case, the two independent wavelet variables, time and space, become separable. Although separability leads to theoretical simplicity and computational efficiency and is, consequently, desirable in image processing, it is not an adequate characteristic for the processing of seismic data.

The signal that we wish to retrieve in seismic data is located along some curve. For instance, for prestack data in various shot/receiver configurations, signals are often assumed to follow a hyperbolic trajectory [Yilmaz1987]. In fact, this trajectory may be more complicated due to inhomogeneities, anisotropy etc., but a hyperbolic assumption is often adequate. Typically, images that are processed by means of wavelet decomposition do not allow such a description. It is the object of PWFD to construct a 2D wavelet based on the hyperbolic signal behavior. Since the independent variables, in this case, are not separable and orthogonal properties no longer apply, conventional 2D wavelets are not the tool of preference. It turns out that frame theory obviates both the separability and orthogonality assumption and, for these reasons, is the approach called wavelet frame denoising. Another important consideration is the redundancy of the frame that is particularly useful in noise suppression.

For a detailed explanation of the PWFD method see Zhang (2000).

Implementation and examples

The success of the 1D WT denoising algorithm depends on the validity of the assumptions discussed before, i.e. that the calculated thresholds are good estimates of the maximum size of the signal at each scale. As a consequence, the method requires enough wavelet coefficients to obtain good statistical estimates. To achieve better statistics in practice, we do not process the data in a trace by trace basis, but rather we form a long time series with all traces arranged in lexicographic order.

In 1D WT filtering, we use a fast discrete wavelet transform [Press 1992] which requires to have a number of elements in the input data equal to a power of 2. In signal processing this requirement is often fulfilled by zero padding. For this work, however, we have chosen to pad not with zeroes but with part of the original data. With this procedure the calculated thresholds are better estimates than when zero padding is applied. In this regard, it is also better to use part of the far offset traces than contain no ground roll, as we want the thresholds to characterize no the noise but the signal.

The process flow is as follows:

1. Form a large trace by lexicographic ordering of the data set (a super-trace).
2. Transform this super-trace into the WT domain, and estimate the threshold values τ at every scale.
3. Downweight the coefficients.
4. Transform back to time domain.
5. Apply PWFD filtering.

Figure 2a shows a shot gather with strong ground roll noise. Figure 2b shows the results after 1D WT filtering. Figure 3 (left) is the same as Figure 2a but showing the data in the wavelet domain, i.e. the WT coefficients: Figure 3 (right) shows the same as Figure 2b in the WT domain, i.e. wavelet coefficients after thresholding. It is clear from these figures that surface waves are not completely removed after 1D WT filtering, but they have been depressed and signals can be seen.
Figure 4a shows again the original shot gather and Figure 4b displays the denoising results by the hybrid method, 1D WT + PWFD filtering. After PWFD filtering, surface waves are further attenuated.

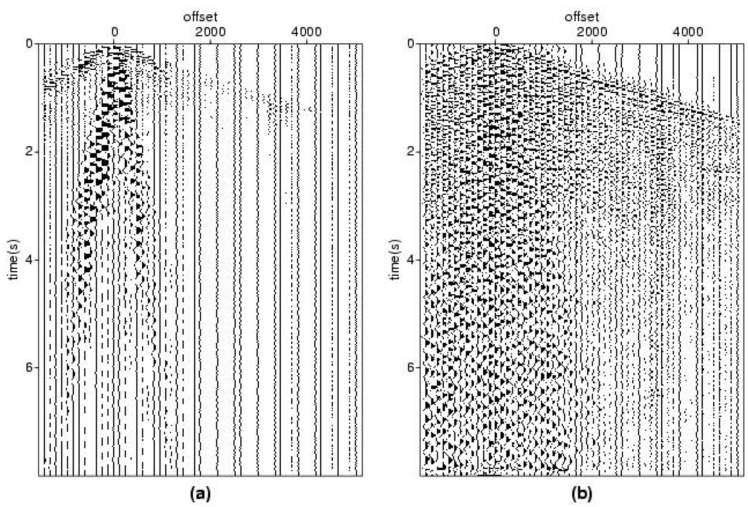

Figure 2: 1D WT processing result. (a) The original shot gather.(b) 1D WT filtering.

Discussion and Conclusions

The PWFD method has proved to be a very powerful and robust tool for noise attenuation. The essential property of this method is the dissimilarity between random noise and a hyperbolic two dimensional wavelet that resembles the nature of signal for most seismic gathers. This quality results in a decomposition of the data where noise maps to coefficients with lower amplitude than those corresponding to signal, allowing us to apply a simple thresholding scheme to eliminate noise. Let us remark that the noise lies in the low amplitude coefficients of the frame domain.

There are situations where noise may map into high amplitude frame coefficients because of its very large energy. The PWFD method may fail in this case. A simple solution to this problem is to apply a complementary filtering with an inverse thresholding, i.e. attenuation of high amplitude coefficients. We have found that the 1D wavelet domain is a good choice for this complementary separation. In this domain ground roll noise and signal have the same similarity with the basis functions, but the high amplitude of the noise makes separation possible.

A very important aspect of the 1D wavelet filtering method is the proper estimation of the thresholds. These estimates determine the success or failure of the method as they define the boundary between signal and noise. We have assumed a Gaussian distribution for the signal and treat the high amplitude coefficients as outliers, which must contain mainly the energy of the ground roll and other interferences. This assumption will be fulfilled as far as we are able to include enough traces without ground roll and the amplitude of this noise is considerably larger than the amplitude of the signal. Note that the noise could have a Gaussian distribution as well, but its variance will be always, for the problem in hand, much larger than the variance of the signal. Because we assume that the bulk of the data is signal, i.e. most traces do not contain high amplitude noise, we can treat this noise as outliers. In this scenario we can apply a robust strategy for outlier removal and attenuate the ground roll. Note that the situation where the method works the best is the same where the PWFD algorithm fails. However, because the noise attenuation is limited we only down weight very large wavelet coefficients and a second pass with the PWFD algorithm is essential to complete the denoising scheme.

We conclude that this hybrid procedure can be very useful for situations where many other methods may fail, mainly because of its complementary nature and robustness.

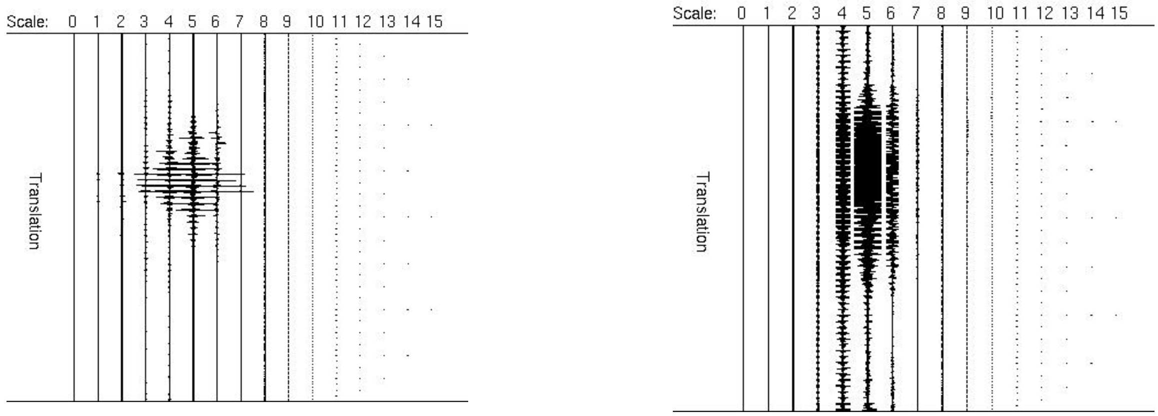

Figure 3: WT coefficients for the super trace: before (left) and after (right) thresholding. The number of coefficients decreases with the increase of scale. (the amplitudes in both plots are scaled by a constant factor for plotting purpose

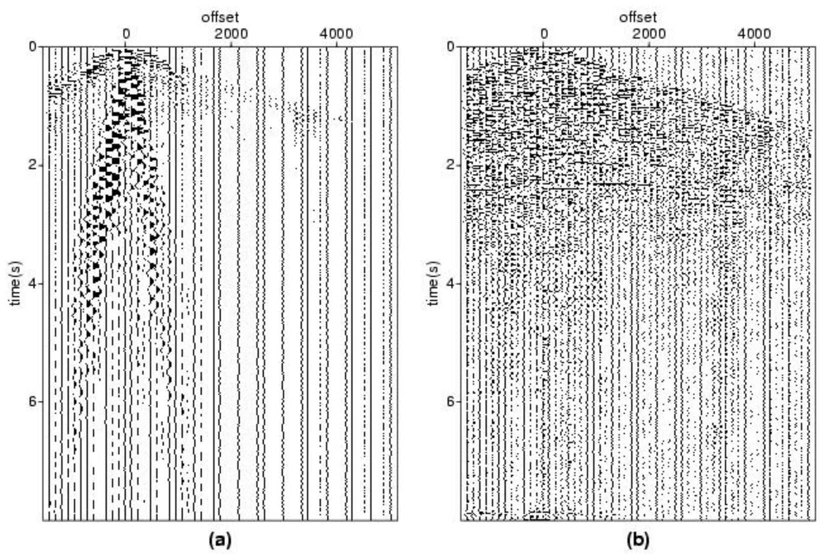

Figure 4: The hybrid approach: (a) The original shot gather.(b) 1D WT + PWFD filtering result.

References

Chakraborty, A., and Okaya, D., 1995, Frequency-time decomposition of seismic data using wavelet-based methods: Geophysics, 60, 1906-1916.

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

Deighan, A. J., and Watts, D. R., 1997, Ground-roll suppression using the wavelet transform: Geophysics, 62, 1896-1903.

Donoho, D. L., 1993, Wavelet shrinkage and w. v. a.- a ten minute tour:, in Meyer, Y. Roques, S., Eds., Progress in wavelet analysis and applications Editions Frontieres, 109-128.

Miao, X., and Cheadle, S. P., 1998, Noise attenuation with wavelet transforms: 68th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1072-1075.

Nguyen, M.-Q., and Mars, J., 1999, Filtering surface waves using 2-D discrete wavelet transform: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1228-1230.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., 1992, Numerical recipes: Cambridge University Press.

Rousseeuw, P. J., and Leroy, A. M., 1987, Robust regression and outlier detection: John Wiley and Sons, Inc.

Trad, D. O., and Travassos, J. M., 2000, Wavelet filtering of magnetotelluric data: Geophysics, 65, 482-491.

Vetterli, M., and Kovacevic, J., 1995, Wavelets and subband coding: Prentice Hall.

Yilmaz, O., 1987, Seismic data processing: Soc. of Expl. Geophys.

Zhang, R., 2000, Physical wavelet frame denoising: CDSST annual report, 171-182.