GEOPHYSICS, VOL. 73, NO. 5 (SEPTEMBER-OCTOBER 2008); P. V37–V46, 9 FIGS.
10.1190/1.2965028

Robust wavelet estimation and blind deconvolution of noisy surface seismics

Mirko Van der Baan[1] and Dinh-Tuan Pham[2]

ABSTRACT

Robust blind deconvolution is a challenging problem, particularly if the bandwidth of the seismic wavelet is narrow to very narrow; that is, if the wavelet bandwidth is similar to its principal frequency. The main problem is to estimate the phase of the wavelet with sufficient accuracy. The mutual information rate is a general-purpose criterion to measure whiteness using statistics of all orders. We modified this criterion to measure robustly the amplitude and phase spectrum of the wavelet in the presence of noise. No minimum phase assumptions were made. After wavelet estimation, we obtained an optimal deconvolution output using Wiener filtering. The new procedure performs well, even for very band-limited data; and it produces frequency-dependent phase estimates.

INTRODUCTION

The objective of blind deconvolution in exploration seismology is to retrieve the reflectivity series without knowledge of the amplitude or phase spectrum of the wavelet. Blind deconvolution differs from conventional deconvolution in that we do not make a priori assumptions about the wavelet phase, which is done routinely in, e.g., predictive deconvolution (gapped/spiking) or spectral whitening.

Wiggins (1978) introduced the first blind deconvolution algorithm based on kurtosis maximization. Convolving any white reflectivity series with an arbitrary wavelet renders the outcome less white but also more Gaussian. Because the earth's reflectivity series is non-Gaussian, which well-log analyses confirm (Walden and Hosken, 1986), the idea is to construct the deconvolution filter so that the output is maximally non-Gaussian (Donoho, 1981).

Maximization of the kurtosis recovers the original reflectivity series because the kurtosis measures the deviation from Gaussianity. The technique can handle nonminimum phase wavelets because the kurtosis is a fourth-order statistic. Higher-order statistics retain phase information, contrary to conventional algorithms based on second-order statistics, such as predictive deconvolution and spectral whitening. Therefore, phase estimation based on evaluation of higher-order statistics is justified theoretically.

Wiggins's algorithm and variants attracted significant attention until the mid-1980s, when it was found that they have several shortcomings. They tend to emphasize the largest reflector at the expense of all others, and they are unstable for very band-limited data (Wiggins, 1985; Longbottom et al., 1988). In particular, if the principal frequency is larger than the wavelet passband (i.e., approximately less than 1.5 octaves of bandwidth), the Wiggins algorithm breaks down (Longbottom et al., 1988; White, 1988).

Now industry practice emphasizes controlled-phase acquisition and processing strategies based predominantly on deterministic corrections (Trantham, 1994). Statistical information contained in the data is used to a much smaller extent these days because it is judged to be less reliable. With the advent of the microelectromechanical system (MEMS) sensors, bandwidth is much less of a problem now than it was 20 years ago; and this supports revisiting the ideas behind the original blind deconvolution algorithms. The algorithms hold the particular promise of estimating the phase of seismic wavelets directly from the data without recourse to seismic-to-well ties.

Any successful deconvolution and wavelet-estimation procedure should be robust in the presence of noise, however, and able to handle band-limited data. We present such a technique based on a modification of the mutual information rate. This rate is a general-purpose statistic to measure whiteness that appeals to higher-order statistics. It is thus capable of estimating the wavelet phase. The developed deconvolution approach uses a two-step procedure. First we estimate the wavelet. Then we deconvolve it from the observed traces using Wiener filtering.

The theoretical derivation of the optimization criterion used for wavelet estimation is presented in Pham (2007). In this paper, we give a simplified, yet self-contained, derivation. Next we focus on the practical implementation of wavelet estimation and blind decon-

Manuscript received by the Editor 17 July 2008; revised manuscript received 4 April 2008; published online 9 September 2008.
[1]University of Leeds, School of Earth and Environment: Earth Sciences, Leeds, United Kingdom. E-mail: m.van-der-baan@see.leeds.ac.uk.
[2]CNRS/INPG, Laboratoire Jean Kuntzman, Grenoble, France. E-mail: Dinh-Tuan.Pham@imag.fr.

volution. Finally, we show some applications on realistic synthetics and a real data example.

THEORY

In the deconvolution problem, we assume that the observed trace x is the result of the convolution of the wavelet w with the reflectivity series r plus some superposed noise n. That is,

$$x(t) = w(t) \star r(t) + n(t), \tag{1}$$

where $\star$ indicates convolution and t represents time. We assume that the reflectivity r and noise n are white, and that the variance of the reflectivity series can be supposed to be one without loss of generality.

The objective is to find a filter $g(t)$ so that the outcome $y(t)$ is as close as possible to the original reflectivity series $r(t)$. Thus, $y(t) = g(t) \star x(t) \approx r(t)$. Because of the presence of noise, the reflectivity series r cannot be recovered perfectly; and a compromise needs to be achieved between noise amplification and successful recovery.

If the wavelet $w(t)$ is known, the Wiener filter $g_w(t)$ achieves the optimum trade-off solution in a least-squares sense (Berkhout, 1977). In the frequency domain, it is given by

$$G_w(f) = \frac{W^*(f)}{|W(f)|^2 + \sigma_n^2}, \tag{2}$$

with f indicating frequency, σ_n^2 indicating the noise variance, and $*$ indicating complex conjugate. Here, and subsequently, we use the convention that capital letters indicate variables in the frequency domain. Hence, Wiener filter G_w is the frequency-domain representation of the time-domain filter g_w.

The mutual information rate is the optimum deconvolution criterion for noiseless data (Pham, 2005). It estimates the whiteness of the deconvolution outcome y using statistics of all orders (Cover and Thomas, 1991). Unfortunately, for noisy band-limited data it boosts the noise outside the passband of the wavelet, which leads to unacceptable results (Larue et al., 2005; Larue and Pham, 2006).

The same problem occurs for conventional deconvolution techniques based on second-order statistics (e.g., predictive deconvolution and spectral whitening). Processing specialists appeal in practice to band-pass filtering after deconvolution to remove the unwanted frequency components. This process is effective but inefficient, and it leaves the subjective choice of what are signal and noise to the interpreter. A better alternative is to estimate the wavelet within its passband and apply Wiener filtering thereafter.

We present a suitable optimization criterion based on a modification of the mutual information rate. This criterion was derived originally in Pham (2007). In the following, we show a simplified but self-contained derivation. We assume continuous variables in the derivation for notational simplicity. Implementation details are described in the appendix.

It can be shown that the mutual information rate C_{mir} can be decomposed as

$$C_{\text{mir}}(y) = -H_{\text{neg}}(y) + \frac{1}{2}\int \log \frac{\sigma_y^2}{S_{yy}(f)} df + \text{cnst}, \tag{3}$$

with $H_{\text{neg}}(y)$ indicating the negentropy, S_{yy} indicating the power spectral density of the outcome y, σ_y^2 indicating its variance, and cnst indicating some constant independent of outcome y (Larue and Pham, 2006; Pham, 2007). In theory, the integration limits run from minus infinity to infinity; but in practice the signal is always sampled, and the spectrum will be wrapped around the Nyquist frequency. As is customary, we take the sampling frequency to be the unit frequency so that the limits reduce to $-1/2$ through $1/2$; that is, the normalized Nyquist frequency. This convention is used in all subsequent integrals.

The negentropy H_{neg}, or standardized negative Shannon entropy, is defined as the difference in differential entropy between a Gaussian variable and a random variable with identical variance (Hyvärinen et al., 2001, sec. 5.4). Gaussian variables have the largest possible differential entropy for a given variance. The negentropy thus is guaranteed to be nonnegative, and it measures the non-Gaussianity of the outcome using statistics of all orders. It can be interpreted as a generalization of the kurtosis. The latter is based on fourth-order statistics. Higher-order statistics have the advantage over second-order ones, such as autocorrelations, because they retain phase information (Mendel, 1991).

The first term in the right-hand side of equation 3 is indispensable because it can identify the phase of the wavelet, whereas the second term invokes second-order statistics and contains no phase information. The second term is related to the second-order whiteness of the outcome. It approaches zero if the power spectral density S_{yy} becomes constant for all frequencies, thus enforcing the outcome y to be white also outside the passband of the wavelet w. This amplifies the noise. Note that both terms in expression 3 are scale invariant. That is, $C_{\text{mir}}(y) = C_{\text{mir}}(\alpha y)$ with α indicating an arbitrary nonzero constant.

The second term is problematic because it is very noise prone. A better alternative is to match the measured power spectral density $\hat{S}_{xx}$ of the observation x to a predicted one, given an estimated wavelet $\hat{w}$ and noise level $\hat{\sigma}_n$. It follows from equation 1 that the theoretical power spectral density S_{xx} is given by

$$S_{xx}(f) = |W(f)|^2 + \sigma_n^2, \tag{4}$$

because both the reflectivity and noise are assumed to be white; and the variance of the reflectivity series can be supposed to be one without loss of generality. In other words, the problem of finding an optimal filter g, so that $C_{\text{mir}}(y = g\text{neg} \star x)$ is minimal, can be changed into the problem of estimating a wavelet $\hat{W}$ and noise level $\hat{\sigma}_n$; so that the estimated power spectral density $\tilde{S}_{xx} = |\hat{W}(f)|^2 + \hat{\sigma}_n^2$, based on model 4, is equal to, or at least proportional to, the measured spectrum $\hat{S}_{xx}(f)$.

The second term in equation 3 can be written as

$$\int \log \frac{\sigma_y^2}{S_{yy}(f)} df = \log\left[\int S_{yy}(f) df\right] - \int \log[S_{yy}(f)] df, \tag{5}$$

where we used the relation $\sigma_y^2 = \int S_{yy} df$. Expression 5 measures the nonflatness of the power spectral density S_{yy}. We want to ensure that the ratio $\hat{S}_{xx}/\tilde{S}_{xx}$ is nearly constant as a function of the frequency.

We consider therefore a similar term in which S_{yy} in equation 3 is replaced by $\hat{S}_{xx}/\tilde{S}_{xx} = \hat{S}_{xx}(|\hat{W}|^2 + \hat{\sigma}_n^2)^{-1}$. If combined with expression 5, this produces

$$C(\hat{W},\hat{\sigma}_n) = -H_{\text{neg}}(y_w) + \frac{1}{2}\log\int\frac{\hat{S}_{xx}(f)}{|\hat{W}(f)|^2+\hat{\sigma}_n^2}df - \frac{1}{2}\int\log\frac{\hat{S}_{xx}(f)}{|\hat{W}(f)|^2+\hat{\sigma}_n^2}df + \text{cnst}, \quad (6)$$

with cnst again indicating some constant independent of the outcome y. The deconvolution outcome y_w is obtained here by Wiener filtering using the estimated wavelet $\hat{w}$ and noise level $\hat{\sigma}_n$.

The optimization criterion 6 depends on the estimated wavelet $\hat{W}$ and noise variance $\hat{\sigma}_n^2$. A joint estimation of the wavelet $\hat{W}$ and noise variance $\hat{\sigma}_n^2$ is difficult; because the first and second terms in the mutual information rate, equation 3, unfortunately are scale invariant. We can determine the power spectral density based on the model $\tilde{S}_{xx} = |\hat{W}|^2 + \hat{\sigma}_n^2$ only as high as a multiplicative constant.

The estimated power spectral density $\tilde{S}_{xx}$ also can be written as

$$\tilde{S}_{xx}(f) = \hat{\sigma}_n^2[|\hat{W}'(f)|^2 + 1], \quad (7)$$

with $\hat{W}' = \hat{W}/\hat{\sigma}_n$, the estimated wavelet divided by the noise standard deviation. The frequency-dependent sum $|\hat{W}'|^2 + 1$ can be estimated for a band-limited wavelet because its spectrum equals one outside the passband of the wavelet, thus providing an absolute reference value.

Substituting the rescaled wavelet $\hat{W}'$ into expression 6 leads to the new optimization criterion that depends on the rescaled wavelet $\hat{W}'$ alone. It is given by

$$C(\hat{W}') = -H_{\text{neg}}(y'_w) + \frac{1}{2}\log\int\frac{\tilde{S}_{xx}(f)}{|\hat{W}'(f)|^2+1}df + \frac{1}{2}\int\log[|\hat{W}'(f)|^2+1]df + \text{cnst}. \quad (8)$$

The associated noise level then can be obtained by inverting expression 7; replacing the estimated power spectral density $\tilde{S}_{xx}$ by the measured one, $\hat{S}_{xx}$; and integrating the result over all frequencies. This yields

$$\hat{\sigma}_n^2 = \int\frac{\hat{S}_{xx}(f)}{|\hat{W}'(f)|^2+1}df. \quad (9)$$

The rescaled output $y'_w = g'_w \star x$ in optimization criterion 8 is estimated using the rescaled Wiener filter G'_w, given in the frequency domain by

$$G'_w(f) = \frac{[\hat{W}'(f)]^*}{|\hat{W}'(f)|^2 + d_n^2}. \quad (10)$$

The prewhitening factor d_n^2 in this expression is put at $d_n = 1$ for conventional Wiener filtering; but it can be increased for improved damping of noisy frequencies, although at the expense of losing some desired signal.

Comparison of expressions 2 and 10 shows that the conventional and rescaled Wiener filters are related by $G_w = G'_w/\hat{\sigma}_n$. Thus, the recovered reflectivity y' might be rescaled to have unit-sample variance by dividing it by the estimated standard deviation $\hat{\sigma}_n$ of the noise.

The combination of wavelet estimation and deconvolution thus starts with an initial guess of the amplitude spectrum of the wavelet, any available phase information, and an initial estimate of the noise variance. The initial wavelet amplitude spectrum is set as the average amplitude spectrum of all observed traces, and the initial noise variance is set to that of the high-frequency components in the data. Next we compute the rescaled Wiener filter 10 and the associated deconvolution outcome y'. A new estimate of the wavelet then is computed by local minimization of criterion 8. This updating cycle is repeated until convergence.

The rationale of the new optimization criterion, expression 8, is that we estimate the frequency-dependent phase of the wavelet by evaluating higher-order statistics. We simultaneously ensure that its amplitude spectrum remains matched to that of the observations. This enhances the robustness of the wavelet estimate because estimation variance increases with the statistical order that is used (Mendel, 1991).

The amplitude spectrum thus is estimated using second-order statistics and only the phase information by means of higher-order statistics. This is in strong contrast to the original blind deconvolution algorithm developed by Wiggins (1978), which derives both the amplitude and phase spectrum of the deconvolution filter from fourth-order statistics. A further advantage of the developed algorithm is that it also yields the estimated wavelet, which can serve for quality-control purposes. A detailed description of the implementation can be found in the appendix.

EXAMPLES

Data set 1: Synthetic example — Noiseless case

To illustrate the performance of the new optimization criterion, two realistic noiseless synthetic gathers are created with the same reflectivity series but with two different source wavelets. The first has a simple 80-degrees-constant-phase rotation with a frequency range between 9 and 32 Hz (i.e., 1.8 octaves and a ratio bandwidth through peak frequency of 1.1). The second wavelet has a complex frequency-dependent phase and a much smaller bandwidth ranging between 15 and 30 Hz (i.e., 1 octave and a ratio bandwidth through peak frequency of 0.67). All quoted frequencies correspond to the −5 db point below the peak frequency.

Estimated wavelets and deconvolution results are compared with those derived using the technique of Levy and Oldenburg (1987), Longbottom et al. (1988), and White (1988), based on kurtosis maximization by constant-phase rotation as implemented by Van der Baan (2008). This method first determines a zero-phase wavelet by spectral averaging of the observed data and then applies a series of constant-phase rotations using different test angles.

The correct phase angle corresponds to the one for which the kurtosis is maximum (Levy and Oldenburg, 1987; Longbottom et al., 1988; White, 1988). The implementation used in Van der Baan (2008) varies slightly from the original method in that it first estimates the source wavelet and then deconvolves it from the original data by means of Wiener filtering. Inspection of the estimated wavelet is useful for quality-control purposes.

Figure 1a displays the noiseless synthetic gather in the time domain for the first simple test wavelet with a constant phase of 80 degrees. Inspection of the waveform of the first arrival shows that it is asymmetrical and therefore nonzero phase. It has equally large positive (black) and negative (white) lobes. Figure 1b shows the deconvolution outcome using Wiener filtering, using the wavelet esti-

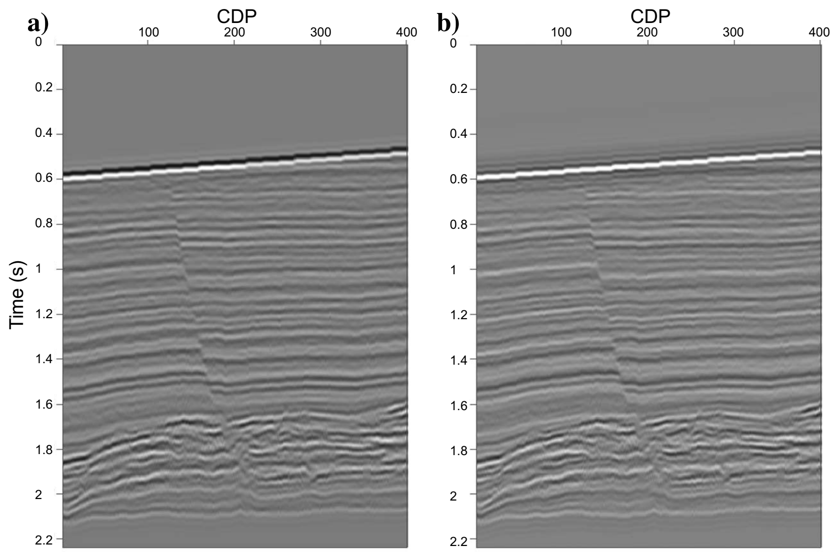

Figure 1. Synthetic example for a simple nonminimum phase wavelet. (a) Original section. (b) Resulting section after deconvolution, which is rendered zero phase. This is most visible around the first arrival, which has a symmetrical waveform after deconvolution.

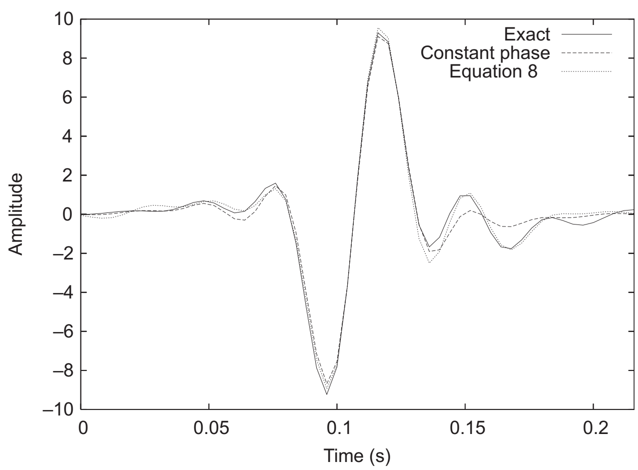

Figure 2. Synthetic example for a simple nonminimum phase wavelet. Both wavelets estimated by constant-phase rotation and with optimization criterion 8 are highly identical to the true wavelet.

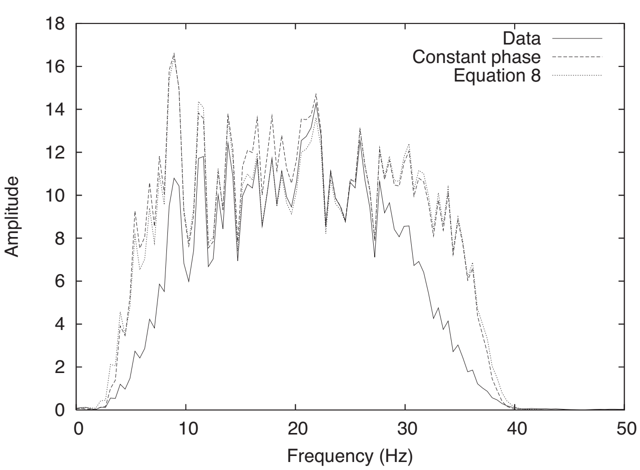

Figure 3. Synthetic example for a simple nonminimum phase wavelet. Deconvolution improves slightly the bandwidth of the data.

mated by minimization of the new optimization criterion, equation 8. The first arrival after deconvolution is symmetrical. It has small positive side lobes with a main negative lobe. The derived optimization criterion led to correct identification of the phase of the source wavelet and produced a zero-phase section.

The simpler method of phase estimation by kurtosis maximization (Levy and Oldenburg, 1987; Longbottom et al., 1988; White, 1988) produces a near-identical wavelet estimate and deconvolution outcome. This is confirmed in Figure 2, which displays the true wavelet and the wavelets estimated by constant-phase kurtosis maximization, and the new frequency-dependent optimization criterion. All results are highly identical as expected for this simple wavelet, although the new criterion 8 produces a slightly better wavelet estimate, especially toward the tail. The slight differences in estimated wavelets have, however, a negligible influence on the deconvolved sections (not shown).

In addition, deconvolution generally improves the bandwidth. This can be seen in Figure 3, which depicts the averaged spectra of all traces before and after deconvolution. The algorithm clearly has improved the bandwidth from 1.8 octaves (9–32 Hz) through 2.3 octaves (7–35 Hz).

The noiseless synthetic section for the complex frequency-dependent narrowband wavelet of 1 octave is shown in Figure 4a; and the deconvolution outcomes are shown in Figure 4b and c, respectively. The ringy nature of the original section, in particular around the first arrival, clearly attests that this wavelet is significantly longer and more complex than the simple constant-phase wavelet used before (compare Figures 1a and 4a). Its bandwidth also is much smaller.

Estimation of the second wavelet is a challenging test because the ratio bandwidth to peak frequency of the wavelet is less than one. Longbottom et al. (1988) and White (1988) demonstrate that kurtosis-based methods will fail for such narrowband wavelets. Indeed, Figure 4b displays the deconvolved section using the wavelet estimated by using constant-phase rotation and kurtosis maximization (Levy and Oldenburg, 1987; Longbottom et al., 1988; White, 1988). The first arrival in the associated deconvolved section remains ringy and asymmetrical (Figure 4b).

Figure 4c shows the deconvolved section after wavelet estimation using optimization criterion 8. The resulting section is simplified greatly and resembles Figure 1b, but with reversed polarity. The wavelet of the first arrival has been compacted greatly and made zero phase, attesting to successful phase estimation.

Figure 5 displays the true and estimated wavelets for this second challenging test. The constant-phase estimate mimics the main lobes of the true wavelet quite well, but it fails to identify any of the side lobes toward its tails. The wavelet estimated by means of optimization criterion 8 estimates both the main and side lobes reasonably well, although some small discrepancies remain visible. Note that the true wavelet has a 0.45-s duration, whereas the entire section is only 2.2 s long.

Figure 6 finally shows the averaged amplitude spectra before and after deconvolution. Both deconvolution results have improved the bandwidth from 1 octave (15–30 Hz) through approximately 1.8 octaves (9–33 Hz); yet only optimization criterion 8 is capable of estimating the correct phase spectrum of the source wavelet (Figure 5).

Data set 1: Synthetic gathers — Noise tests

The previous wavelet estimates are obtained using noiseless data; therefore they are obtained under optimal conditions. They demonstrate the best performance possible for both wavelet estimation techniques. In the following, we explore their robustness in the presence of noise. We use the same two synthetic gathers as before, but we add random white Gaussian noise using a range of signal-to-noise ratios and compare performances.

The correlation coefficients between the true and estimated wavelets are computed for 50 Monte Carlo simulations per noise test, and their averages and standard deviations are calculated. Different noise realizations are imposed on each trace; yet the signal-to-noise ratio remains fixed in each test. All signal-to-noise ratios are defined as the standard deviation of the signal over that of the noise.

Correlation coefficients have a maximum value of one, which biases estimated standard deviations if unaccounted for. We apply the Fisher transformation on all computed values. This process transforms the correlation coefficients to an approximate Gaussian distribution. Then we compute the average and the positive and negative

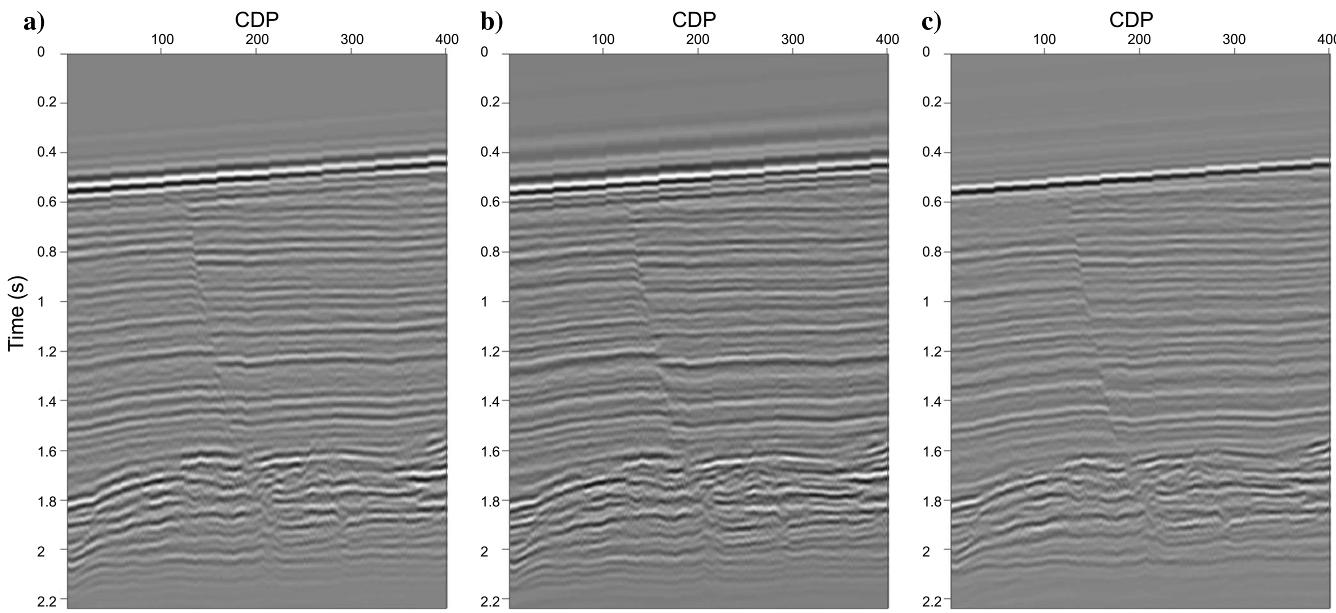

Figure 4. Synthetic example for a complex nonminimum-phase wavelet. (a) Original section. (b) Deconvolution result using a constant-phase wavelet. (c) Deconvolution result using a frequency-dependent wavelet estimated by using optimization criterion 8. Only the frequency-dependent wavelet renders the time-domain output zero phase.

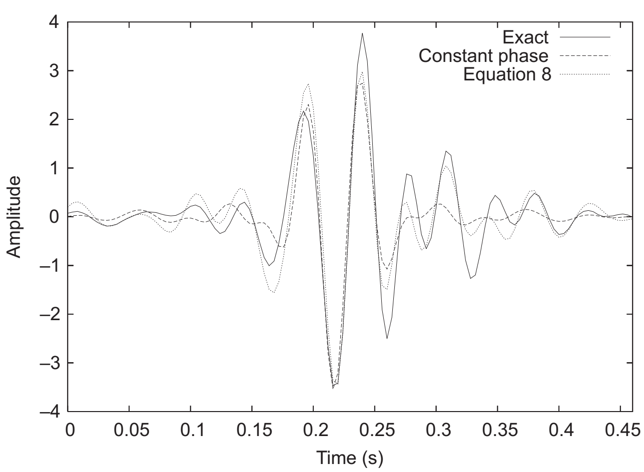

Figure 5. Synthetic example for a complex nonminimum-phase wavelet. Only the wavelet derived using the modified mutual-information-rate criterion 8 derives a reasonably close estimate to the true wavelet. The one obtained by constant-phase rotation recovers only the main lobes.

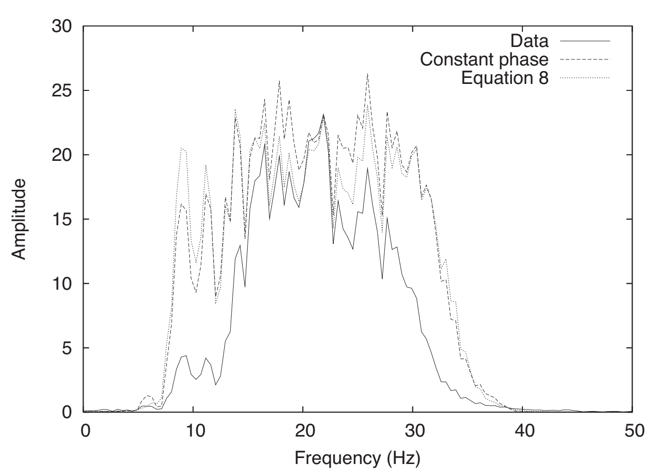

Figure 6. Synthetic example for a simple nonminimum-phase wavelet. Both deconvolution results improve the bandwidth of the data, despite the fact that only optimization criterion 8 correctly identifies the phase of the wavelet.

standard deviation. Finally, we transform the inverse transform to these values. See Vandecar and Crosson (1990) for further details.

Figure 7 contains the correlation coefficients found by using both techniques for both wavelets and various noise levels. Both wavelet-estimation techniques perform well for high signal-to-noise ratios for the simple wavelet (Figure 7a). The constant-phase rotation method performs best. It achieves, in general, better reconstructions, as seen by the higher correlation coefficients; and its standard deviations are significantly lower at all noise levels. The simpler wavelet estimation technique outperforms results obtained by minimization of optimization criterion 8. This is because the latter involves more inversion parameters and therefore more degrees of freedom, which are redundant for this simple wavelet, thereby enhancing proneness to noise contamination.

The complex wavelet cannot be described by a constant phase, and wavelet estimation using constant-phase estimation therefore fails for all signal-to-noise ratios (Figure 7b). Estimated correlation coefficients in all cases are below 0.92. Optimization criterion 8, on the other hand, performs remarkably well, even for very low signal-to-noise levels of 0.25, and despite the limited bandwidth of the data.

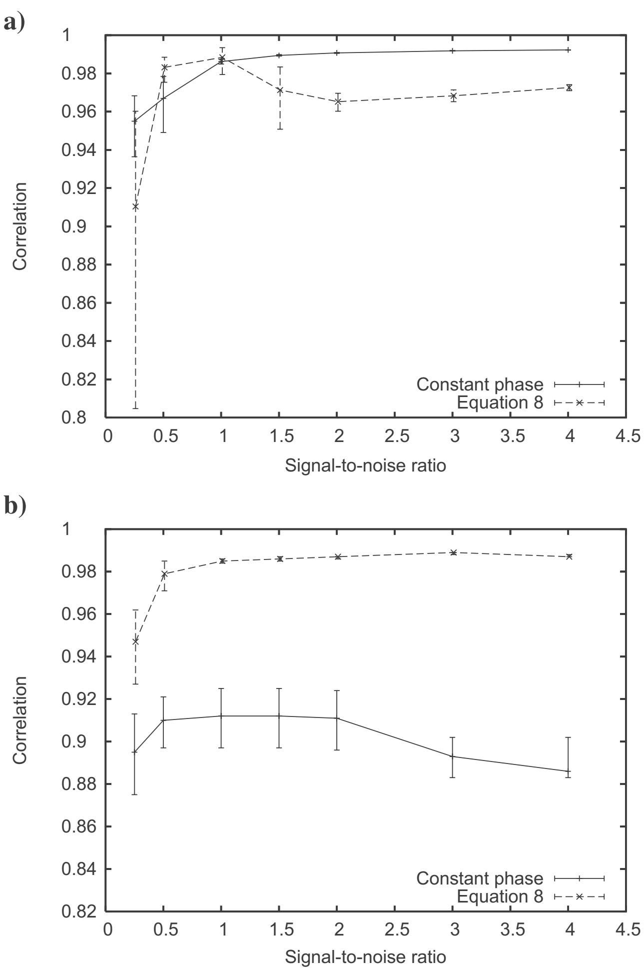

Figure 7. Noise stability tests for both synthetic examples. Correlation coefficients between estimated and true wavelet as a function of signal-to-noise ratio for (a) simple constant-phase wavelet, and (b) complex wavelet with frequency-dependent phase.

Of course, the large number of traces (400) helps to enhance performance; yet the developed wavelet-estimation technique is clearly robust in the presence of noise.

Data set 2: Stacked section

Finally, we apply the new wavelet estimation and deconvolution approach on a real section. The results before and after deconvolution are shown in Figure 8. The clearest changes occur around the ocean bottom at 0.55 s. The deconvolved section depicts a much simpler reflectivity here. It also emphasizes the undulating interface about 1.6 s. This can be seen best in Figure 8c and d, respectively, which zoom in on the ocean-bottom reflection before and after deconvolution.

Figure 9 contains the wavelets estimated by using both constant-phase kurtosis maximization and the new optimization criterion. Estimates agree, but they produce surprisingly complex wavelets for this stacked section — in particular, the wavelet is not zero phase, and it has a ringy character. This attests to the importance of checking the phase of the wavelet as a quality control in the last processing steps of a data set. Both wavelets produce near-identical deconvolved sections (not shown).

DISCUSSION

Broadly speaking, most deconvolution criteria can be divided into two categories. They maximize either the whiteness or the non-Gaussianity of the deconvolution outcome. The whiteness criteria are based on the fact that convolution of any filter with a white reflectivity series renders the outcome less white. The correct deconvolution filter therefore is the one that produces the whitest deconvolution result. Conversely, non-Gaussianity criteria make use of the fact that convolution of any filter with a white reflectivity series renders the outcome more Gaussian. The correct deconvolution filter is the one that leads to the most non-Gaussian deconvolution result (Donoho, 1981).

In any case, both deconvolution criteria based on whiteness or non-Gaussianity considerations must appeal to higher-order statistics to determine the phase of the wavelet; because second-order statistics, such as autocorrelation functions, reveal no phase information (Mendel, 1991). This assumes, of course, that the earth has a non-Gaussian reflectivity series, which is a fact supported by well-log analyses (Walden and Hosken, 1986).

Deconvolution criteria based on whiteness have the disadvantage of boosting the noise level outside the passband of the wavelet because they enforce whiteness over the entire frequency range. This leads to problems if the actual wavelet is band limited. Processing specialists thus appeal to band-pass filtering after deconvolution to eliminate the unwanted frequency components.

Deconvolution criteria based on non-Gaussianity, such as the original method proposed by Wiggins (1978), ignore entirely the amplitude spectrum of the observations and evaluate solely whether a specific filter leads to a more or less Gaussian deconvolution result. The constant-phase rotation approach of Levy and Oldenburg (1987), Longbottom et al. (1988), and White (1988) also falls into this category.

The hybrid optimization criterion 8 of Pham (2007) for robust wavelet estimation is unusual, because it evaluates the non-Gaussianity of the deconvolution result and yet ensures that the estimated wavelet spectrum remains matched to the amplitude spectrum of the observations. The first term in the right-hand side of equation 8 thus determines the phase of the wavelet by evaluating the negentropy, a general-purpose non-Gaussianity criterion that appeals to statistics of all orders. The second and third terms guarantee that the wavelet has the correct amplitude spectrum.

The modified mutual-information-rate criterion 8 therefore is designed to be robust in the presence of noise, and it is capable of handling band-limited wavelets. Nonetheless, it remains possible that high-frequency noise would leak into the wavelet estimate. If this is suspected, use the usual remedies of (1) band-pass filtering the wavelet prior to Wiener filtering, (2) band-pass filtering the deconvolution output, and/or (3) increasing the prewhitening factor d_n in expression 10. It is also possible to use the wavelet estimated by constant-phase kurtosis maximization as an initial guess for frequency-dependent phase estimation using optimization criterion 8. This will also render inversion results more stable and improve convergence.

We envision that estimation criterion 8 is most useful for very band-limited data (e.g., old legacy data) and situations in which long wavelets with complex phases are anticipated (e.g., a remnant vibroseis or a dispersive wavelet). Successful phase estimation by means of a statistical analysis of the data produces also an independent estimate from the one obtained using seismic-to-well ties. This method can be used as a quality control to assess whether or not processing has introduced unexpected phase changes.

Finally, the method assumes that the wavelet remains unchanged over the entire analyzed section. If nonstationary wavelets are anticipated, e.g., as a result of the presence of attenuation, time-varying wavelet estimation and deconvolution are required. A suitable technique is developed in Van der Baan (2008), who extended the constant-phase method of Levy and Oldenburg (1987), Longbottom et al. (1988), and White (1988) to the nonstationary case.

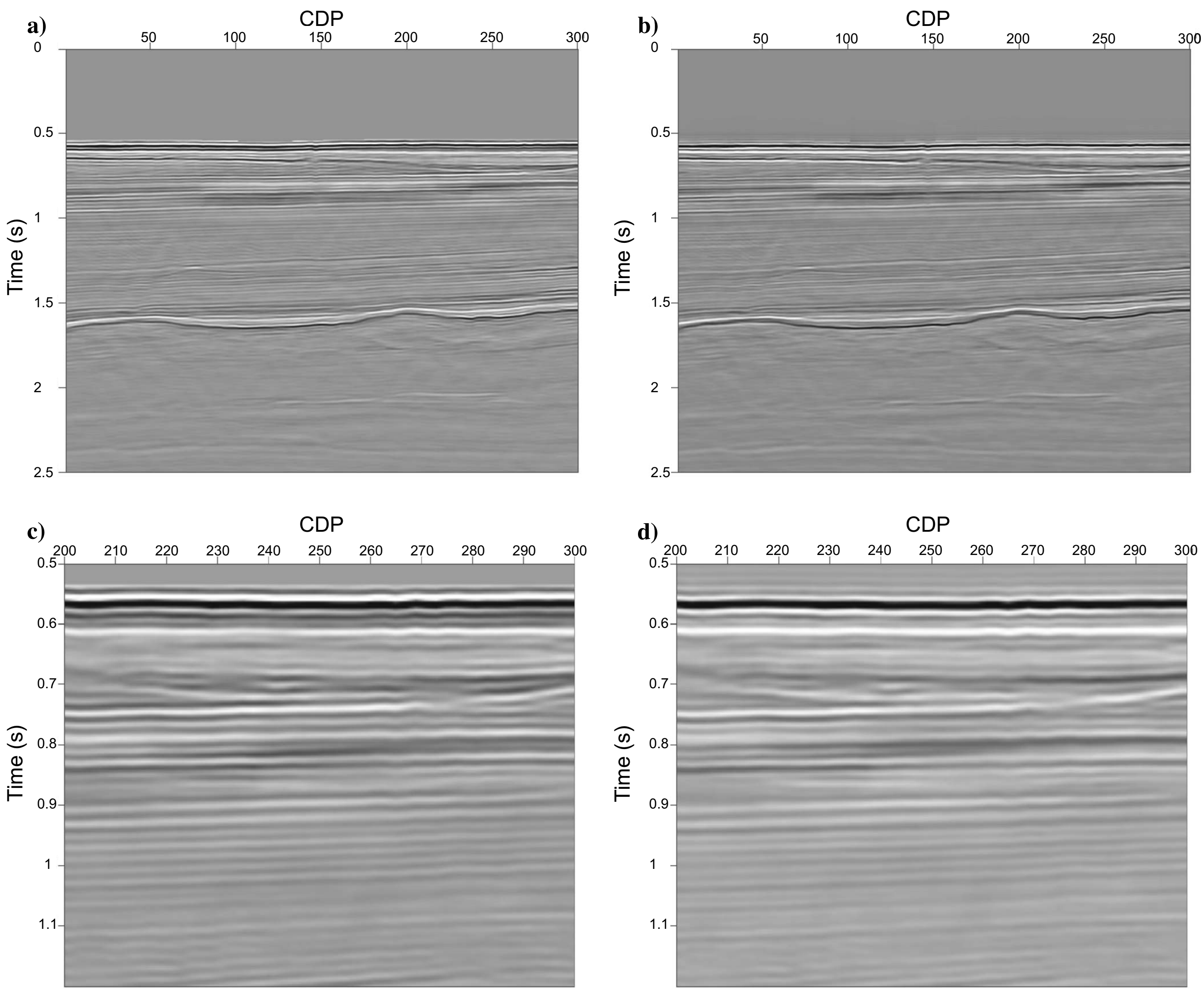

Figure 8. Real stacked section. (a) Before deconvolution, and (b) after deconvolution; (c) and (d) are zoom-ins on the ocean-bottom reflection of, respectively, the original data and the deconvolution results. Deconvolution simplifies the ocean-bottom reflection. Wiener filtering using the wavelets estimated by, respectively, the constant-phase rotation approach and the new optimization criterion lead to similar results in this case.

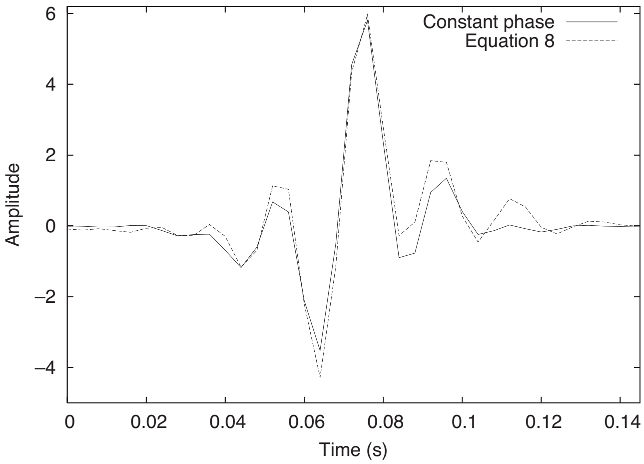

Figure 9. Estimated wavelets for the stacked section displayed in Figure 8. The constant-phase rotation approach and the new optimization criterion lead to similar wavelet estimates in this case.

CONCLUSIONS

Robust blind deconvolution of noisy surface-seismic data is a challenging problem. It is particularly difficult to estimate accurately the phase of the wavelet. Most blind deconvolution algorithms assume theoretically noiseless data and a large wavelet passband, which often leads to poor performance on seismic data. Deconvolution methods based on whiteness criteria boost the noise outside the wavelet passband, whereas approaches based on non-Gaussianity criteria entirely ignore the amplitude spectrum of the observations.

The modified mutual-information-rate criterion leads to a procedure for wavelet estimation and blind deconvolution that is robust in the presence of noise. It is designed specifically to handle band-limited wavelets. The wavelet phase can be estimated using the negentropy, a general-purpose non-Gaussianity criterion. Simultaneously, we ensure that the wavelet amplitude spectrum remains matched to that of the observations.

The resulting two-step inversion technique first estimates the wavelet and then deconvolves it from the observed data using Wiener filtering. The Monte Carlo noise tests on realistic synthetic gathers demonstrate that statistical wavelet estimation is feasible from noisy surface seismics alone, even for very band-limited wavelets with a complex, frequency-dependent phase, under these conditions: (1) the wavelet displays temporal and spatial invariance within the analyzed section, (2) the reflectivity series is white and non-Gaussian, and (3) the superposed noise is white.

ACKNOWLEDGMENTS

The authors thank the BG group, BP, Chevron, the Department of Trade and Industry, and Shell for financial support of the project Blind Identification of Seismic Signals (BLISS). Comments and suggestions by Roald van Borselen, Jinghuai Gao, Eric Verschuur, and an anonymous reviewer helped to improve the original manuscript. We are grateful to Dave Wilkinson for creating the synthetic sections, and to BP and Chevron for permission to use, respectively, the real and synthetic data examples.

APPENDIX A

IMPLEMENTATION

Blind deconvolution is achieved using an iterative two-step procedure. First we estimate the wavelet; then we deconvolve it from the observed traces by means of Wiener filtering, equation 10. Optimization criterion 8 is minimized using a quasi-Newton scheme to yield a new estimate of the wavelet. This process is repeated until convergence.

In this appendix, we describe the most important implementation steps. Note first, though, that the new optimization criterion 8 is of a hybrid nature in the sense that it involves both the time domain wavelet $\hat{w}$ and its frequency representation $\hat{W}$. This hybrid form is chosen because the negentropy and its derivatives are computed most easily in the time domain, but the constraints on amplitude spectra are best evaluated in the frequency domain. We update the desired wavelet in the time domain and compute its frequency representation by a discrete Fourier transform.

Several items are important for the wavelet estimation, namely noise injection, proper initialization, computation of the negentropy and the entropy, spectral estimation, computation of the gradient, and the multitrace implementation.

Noise injection

The optimization criterion 8 assumes that data are contaminated by white noise, equation 4. In reality, this is rarely the case — in particular, because band-pass filtering often is applied prior to deconvolution filter design. To stabilize the method, we appeal to a small amount of noise injection before wavelet estimation to ensure that a minimum level of white noise exists. The energy of the injected noise is recommended to be less than 1% of the signal energy.

The noise is unlikely to be white in practice, thus introducing some bias in the wavelet estimation. This issue and possible remedies are considered further in the Discussion section.

Initialization

We use a zero-phase wavelet as the initial guess in the inversion. Its amplitude spectrum is set equal to the average spectrum of the observed traces. The phase spectrum is zero. Linear tapering is used in the time domain to reduce artifacts at the wavelet edges and to set a priori constraints on the anticipated time duration of the wavelet. Details can be found in Van der Baan (2008).

Alternatively, we could use a minimum-phase initial guess or estimate a frequency-independent phase rotation using the kurtosis (Levy and Oldenburg, 1987; Longbottom et al., 1988; Van der Baan, 2008). A more accurate initialization leads to faster convergence.

Negentropy and entropy computation

The negentropy H_{neg} measures the deviation from a Gaussian distribution using statistics of all orders. It is defined as the difference in entropy between a Gaussian signal and the analyzed signal, when both have the same covariance matrix. It can be decomposed as

$$H_{\text{neg}}(y) = \frac{1}{2}\log(2\pi\sigma_y^2) - H(y), \tag{A-1}$$

with $H(y)$ indicating the differential entropy of outcome y. The first term is the entropy of a Gaussian variable for a given variance. We approximate the entropy $H(y)$ using the kernel estimator of Pham

(2004). The kernel used is a third-order cardinal spline. This method estimates the entropy and its derivatives at low computational cost.

Note that the negentropy is scale invariant. Hence $H_{\mathrm{neg}}(y_w) = H_{\mathrm{neg}}(y'_w)$ because $H(\alpha y) = H(y) + \log|\alpha|$, with α indicating some nonzero constant.

Discrete spectral estimation

The reduced optimization criterion $C(\hat{W}')$, equation 8, relies heavily on correct estimation of the power spectral density $\hat{S}_{xx}$. Estimation of the power spectrum of stochastic processes normally requires some thought and often smoothing or averaging. The integrations involved in the computation of the optimization criterion $C(\hat{W}')$ provide implicit smoothing.

Therefore, we simply estimate the discrete power spectral density $\hat{S}_{xx}$ using the periodogram; that is,

$$\hat{S}_{xx}(f_k) = \frac{1}{n_f} X(f_k) X^*(f_k), \quad \text{(A-2)}$$

with n_f indicating the number of discrete frequencies f_k, and $X(f_k)$ indicating the Fourier transform of observation $x(t)$. The cross-spectra involved in the gradient of the optimization criterion are estimated in similar ways. Using raw, unsmoothed periodograms for spectral estimation also allows us to replace all integrations in the continuous optimization criterion 8 by Riemann summations over all discrete frequencies f_k.

Gradient

Optimization criterion 8 is minimized using a quasi-Newton scheme, which requires the gradient of optimization criterion $C(\hat{W}')$ with respect to the time-domain wavelet amplitudes, but which has a quadratic convergence near its minimum (Press et al., 1992, chap. 10). For completeness, we repeat the gradient here. Its derivation can be found in Pham (2007). It is given by

$$\frac{\partial C(\hat{W}')}{\partial \hat{w}_t} = \frac{1}{n_f} FT^{-1} \left\{ \frac{\hat{S}_{x\phi}(f_k)}{|\hat{W}'(f)|^2 + 1} - 2[G'_w(f_k)]^* \mathrm{Re}[G'_w(f_k)\hat{S}_{x\phi}(f_k)] + [G'_w(f_k)]^* \left[1 - \frac{\hat{S}_{xx}(f_k)}{\hat{\sigma}_n^2(|\hat{W}'(f_k)|^2 + 1)} \right] \right\}, \quad \text{(A-3)}$$

with $\hat{w}_t$ indicating the estimated wavelet at discrete time t, FT^{-1} the inverse Fourier transform, and $\mathrm{Re}[.]$ the real part.

The cross-spectrum $\hat{S}_{x\phi}(f_k)$ is given by

$$\hat{S}_{x\phi}(f_k) = \frac{1}{n_f} \left[X(f_k)\Psi^*(f_k) - \frac{X(f_k)[Y'_w(f_k)]^*}{\sigma^2_{y'}} \right], \quad \text{(A-4)}$$

where $\Psi(f_k)$ is the Fourier transform of the time-domain score function $\psi(t)$ defined by

$$\psi[y'_t] = n_f \frac{\partial H[y'_t]}{\partial y'_t}, \quad \text{(A-5)}$$

with y'_t indicating the rescaled deconvolution output at discrete time t.

The score function ψ is the partial derivative of the entropy H of outcome y. It is estimated again by using the kernel estimator of Pham (2004). After each iteration, the noise variance $\hat{\sigma}_n^2$ is computed by equation 9.

Multitrace implementation

The algorithm has been described for a single-trace implementation. Multitrace implementations are straightforward. We compute the optimization criterion 8 and gradient A-3 for each trace and subsequently average them. The averaged values then are used in the inversion algorithm to update the time-domain wavelet. Likewise, the estimated noise variance $\hat{\sigma}_n^2$ is averaged over all traces. A multitrace implementation significantly enhances the stability of the inversion results because it leads to more robust estimation of the negentropy and its score function.

REFERENCES

Berkhout, A. J., 1977, Least-squares inverse filtering and wavelet deconvolution: Geophysics, **42**, 1369–1383.

Cover, T., and J. Thomas, 1991, Elements of information theory: Wiley.

Donoho, D., 1981, On minimum entropy deconvolution, *in* D. F. Findley, ed., Applied time series analysis II: Academic Press, 565–608.

Hyvärinen, A., J. Karhunen, and E. Oja, 2001, Independent component analysis: Wiley.

Larue, A., and D. T. Pham, 2006, Comparison of supergaussianity and whiteness assumptions for blind deconvolution in noisy context: Presented at the 14th European Signal Processing Conference.

Larue, A., M. Van der Baan, J. I. Mars, and C. Jutten, 2005, Sparsity or whiteness: What criterion to use for blind deconvolution of seismic data?: 75th Annual International Meeting, SEG, Expanded Abstracts, SI 1.4.

Levy, S., and D. W. Oldenburg, 1987, Automatic phase correction of common-midpoint stacked data: Geophysics, **52**, 51–59.

Longbottom, J., A. T. Walden, and R. E. White, 1988, Principles and application of maximum kurtosis phase estimation: Geophysical Prospecting, **36**, 115–138.

Mendel, J. M., 1991, Tutorial on higher-order statistics (spectra) in signal processing and system theory: Proceedings of the IEEE, **79**, 278–305.

Pham, D. T., 2004, Fast algorithms for mutual information-based independent component analysis: IEEE Transaction on Signal Processing, **52**, 2690–2700.

———, 2005, Mutual information approach to blind separation-deconvolution: Presented at the 13th European Signal Processing Conference.

———, 2007, Blind deconvolution in noisy and band-limited context: Presented at the 15th European Signal Processing Conference.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992, Numerical recipes in Fortran: The art of scientific computing: 2nd ed., Cambridge University Press.

Trantham, E. C., 1994, Controlled-phase acquisition and processing: 64th Annual International Meeting, SEG, Expanded Abstracts, 890–894.

Van der Baan, M., 2008, Time-varying wavelet estimation and deconvolu-

tion by kurtosis maximization: Geophysics, **73**, no. 2, V11–V18.
Vandecar, J. C., and R. S. Crosson, 1990, Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least-squares: Bulletin of the Seismological Society of America, **80**, 150–169.
Walden, A. T., and J. W. J. Hosken, 1986, The nature of the non-gaussianity of primary reflection coefficients and its significance for deconvolution: Geophysical Prospecting, **34**, 1038–1066.
White, R. E., 1988, Maximum kurtosis phase correction: Geophysical Journal, **95**, 371–389.
Wiggins, R. A., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21–35.
——–, 1985, Entropy guided deconvolution: Geophysics, **50**, 2720–2726.