Zero- and first-order approximations for least-squares estimation of seismic signal contaminated by coherent and random noise of complex structure

Yuriy Tyapkin, UkrSGPI, BjornUrsin, NTNU, Olena Silinska, UkrSGPI, OlenaTiapkina, NTNU*

Summary

A method for least-squares signal estimation with a complicated and therefore more realistic mathematical model of the multichannel seismic record containing random noise and an arbitrary number of coherent noise wavetrains is described. Under certain conditions, the method may be reduced to two successive stages: preliminary subtraction of estimates of all the coherent noise wavetrains and final estimation of the signal from the residual record. A simplified scheme and an advanced scheme for subtracting coherent noise are proposed, which are called zero-order and first-order approximations respectively. The first of them is the generalization of the conventional scheme for subtracting coherent noise to the complicated data model. The second scheme has indisputable advantages over the first one, since it allows the distortions that appear when estimating and subsequently subtracting the coherent noise wavetrains to be compensated. The effectiveness of the advanced technology is demonstrated with synthetic data sets contaminated by coherent noise of various intensities.

Introduction

When searching for and prospecting of hydrocarbon traps, geophysicists are often faced with the problem of recovering signal from multichannel seismic data sets contaminated by spatially coherent and random noise. Among both kinds of noise, it is coherent noise that is the most persistent problem in seismic imaging, and a number of techniques have been developed to attenuate it: optimum weighted stacking (Schoenberger, 1996), optimum array filtering (Hanna and Simaan, 1985), radial trace filtering (Zhu et al., 2004), *f-k* filtering (March and Bailey, 1983), Radon (τ-p) transform (Mitchell and Kelamis, 1990; Sacchi and Ulrych, 1995), spectral matrix filtering (Gounon et al., 1998), singular value decomposition (Ulrych et al., 1999; Trickett, 2003; Tyapkin et al., 2007). Recently, in order to take advantage of high order statistics, the last technique has been combined with independent component analysis (Vrabie et al., 2004; Bekara and Van der Baan, 2006). In the 1960s-1970s, in order to solve the problem of optimally recovering the seismic signal imbedded in coherent and random noise, the majority of publications exploited multichannel Wiener filters (Meyerhoff, 1966; Cassano and Rocca, 1974,) or maximum-likelihood (least-squares) estimators. Then the situation changed and the interest of geophysicists in these methods gradually was lost because of their insufficient effectiveness. Their place was taken by non-optimum methods implying simplified mathematical models of seismic data, with *f-k* filtering and Radon transform being best known among them. Even though these filters are usually faster and more cost-effective, the most undesirable aspect of them is the mixing of the data, which is usually inherent in these processes, producing a wormy appearance in the output data. This leads to signal distortion and spatial correlation of background noise. In our opinion, the optimum methods are often less effective because they exploit, first, imperfect mathematical models of the record and, second, an imperfect scheme for subtracting coherent noise. For this reason, in this paper, we make an attempt to rehabilitate these methods and to reanimate the interest of geophysicists in them. With this purpose in mind, we utilize a more complicated and adequate mathematical model of the record. We adopt this generalized record model in order to get rid of or at least to minimize the model assumption errors, which refer to deviations of actual records from the assumed model, and thus to obtain more reliable signal estimates. With this mathematical model given, we derive a least-squares method for optimally approximating the underlying signal. Under certain conditions, the method may be reduced to two successive stages: preliminary subtraction of estimates of all the coherent noise wavetrains and final estimation of the signal from the residual record. On both stages, optimum weighted stacking is used with reference to the variance of random noise and the amplitudes and arrival times of the corresponding coherent components. A simplified scheme and an advanced scheme for subtracting coherent noise are proposed, which are called zero-order and first-order approximations respectively. The first of them is the generalization of the conventional scheme for subtracting coherent noise to the complicated data model. The second scheme has indisputable advantages over the first one, since it allows the distortions that appear when estimating and subsequently subtracting the coherent noise wavetrains to be compensated. The effectiveness of the advanced technology is demonstrated with synthetic data sets contaminated by coherent noise of various intensities.

Solution of the problem and its analysis

Suppose that the ith trace of the record that contains N traces may be written as:

$$u_i(t) = a_i s\left(t - \tau_{(s)i}\right) + \sum_{l=1}^{L} b_{il} r_l\left(t - \tau_{(r)il}\right) + n_i(t),\ i = 1, \ldots, N\,. \quad (1)$$

Here the signal is described by the first term on the right-hand side of equation (1) and assumed to have an identical waveform $s(t)$ on each trace, with arbitrary trace-dependent

Zero- and first-order approximations for least squares estimation of seismic signal

amplitudes a_i and time delays $\tau_{(s)i}$. The amplitudes are permitted to have zero (a signal-free trace) and negative (e.g., due to the AVO-effect) values. The second term represents a superposition of coherent noise wavetrains with individual waveforms $r_l(t),\ l = 1, ..., L$. Each of the wavetrains, as well as the signal, bears arbitrary trace-dependent amplitudes b_{il} and time delays $\tau_{(r)il}$. Random noise is expressed by the third term and supposed to be a stationary zero-mean Gaussian stochastic process independent from trace to trace, with identical to within a scale factor, the variance σ_i^2, autocorrelations on different traces. All the coherent noise waveforms are also assumed to be stationary zero-mean Gaussian stochastic processes independent of the signal, random noise and each other. Due to the above assumptions, the cross-spectrum of the entire noise between channels i and j may be expressed as:

$$R_{ij}(\omega) = \sum_{l=1}^{L} b_{il} b_{jl} R_{(r)l}(\omega) \exp\left[i\omega\left(\tau_{(r)jl} - \tau_{(r)il}\right)\right] + \sigma_i^2 R_{(n)}(\omega)\delta_{ij}, \quad (2)$$

where $R_{(r)l}(\omega)$ and $R_{(n)}(\omega)$ are the power spectra of $r_l(t)$ and any $n_i(t)$, respectively, at the angular frequency ω, while δ_{ij} signifies the Kronecker delta function.

Given this model, let us state the problem to obtain the best in least-squares sense estimate of the signal shape, $s(t)$. With this purpose in mind, we represent the signal component as $\mathbf{f}s$, where the column vector $\mathbf{f}$ is written $\mathbf{f} = \{a_1 \exp(i\omega\tau_{(s)1}), ..., a_N \exp(i\omega\tau_{(s)N})\}^H$, the scalar s is the Fourier spectrum of $s(t)$, and the superscript H denotes complex conjugate (Hermitian) transpose. For shot, here and in the following the functional dependence on frequency is dropped. The least-squares estimate of s can be written (Helstrom, 1968)

$$s = \left(\mathbf{f}^H \mathbf{R}^{-1} \mathbf{f}\right)^{-1} \mathbf{f}^H \mathbf{R}^{-1} \mathbf{u}, \quad (3)$$

where the column vector $\mathbf{u} = \{U_1^*, ..., U_N^*\}^H$ contains the Fourier spectra U_i, $i = 1, \ldots, N$, of all the traces, $\mathbf{R}$ is the matrix whose elements are defined by equation (2), -1 signifies matrix inverse, and the superscripted asterisk stands for complex conjugation.

In order to invert $\mathbf{R}$, let us represent this matrix in the form

$$\mathbf{R} = \mathbf{G}\mathbf{B}\mathbf{G}^H + R_{(n)}\mathbf{D}, \quad (4)$$

where $\mathbf{G} = \{\mathbf{g}_j\} = \{b_{ij} \exp(-i\omega\tau_{(r)ij})\},\ i = 1, ..., N,\ j = 1, ..., L$, is an N by L matrix that consists of the column vectors

$$\mathbf{g}_j = \{b_{1j} \exp(i\omega\tau_{(r)1j}), ..., b_{Nj} \exp(i\omega\tau_{(r)Nj})\}^H,$$

$$\mathbf{B} = \mathrm{diag}\{R_{(r)1}, ..., R_{(r)L}\}, \quad \mathbf{D} = \mathrm{diag}\{\sigma_1^2, ..., \sigma_N^2\}.$$

Then we can take advantage of the method described by Horn and Johnson (1986) and obtain

$$\mathbf{R}^{-1} = R_{(n)}^{-1}\left(\mathbf{I} - R_{(n)}^{-1}\mathbf{D}^{-1}\mathbf{G}\mathbf{V}^{-1}\mathbf{G}^H\right)\mathbf{D}^{-1}, \quad (5)$$

where $\mathbf{I}$ is an identity matrix,

$$\mathbf{V} = \mathbf{B}^{-1} + R_{(n)}^{-1}\mathbf{G}^H\mathbf{D}^{-1}\mathbf{G}. \quad (6)$$

Let the inequality

$$R_{(r)i} R_{(n)}^{-1} c_{ii} >> 1, \quad (7)$$

where

$$c_{ij} = \mathbf{g}_i^H \mathbf{D}^{-1} \mathbf{g}_j = \sum_{k=1}^{N} \frac{b_{ki} b_{kj}}{\sigma_k^2} \exp\left[i\omega\left(\tau_{(r)ki} - \tau_{(r)kj}\right)\right],\ i, j = 1, ..., L, \quad (8)$$

is valid for all i. The left-hand side of this inequality is the ratio of the power spectra of the ith coherent noise and the random noise at the output from optimum weighted stacking performed with regard for the amplitudes and arrival times of this coherent noise (Tyapkin and Ursin, 2005). From equation (8), it follows that this quantity may be expressed as:

$$R_{(r)i} R_{(n)}^{-1} c_{ii} = \sum_{k=1}^{N} \frac{R_{(r)i} b_{ki}^2}{R_{(n)}\sigma_k^2} = \sum_{k=1}^{N} R_k^{(i)}, \quad (9)$$

where $R_k^{(i)}$ is the ratio of the above spectra on the kth trace. From this it follows that inequality (7) can be valid even when the value of $R_k^{(i)}$ is rather small on each trace but the number of traces involved in processing is large enough.

If the record obeys inequality (7), we can neglect the first term on the right-hand side of equation (6) and obtain

$$\mathbf{V} = R_{(n)}^{-1}\mathbf{G}^H\mathbf{D}^{-1}\mathbf{G}. \quad (10)$$

This matrix with the elements $R_{(n)}^{-1}c_{ij}$ can be expressed as

$$\mathbf{V} = \mathbf{V}_0 + \mathbf{V}_1 = \mathbf{V}_0\left(\mathbf{I} + \mathbf{V}_0^{-1}\mathbf{V}_1\right) = \mathbf{V}_0(\mathbf{I} + \mathbf{E}), \quad (11)$$

where $\mathbf{V}_0$ and $\mathbf{V}_1$ contain the diagonal and off-diagonal elements of $\mathbf{V}$ respectively, $\mathbf{E} = \mathbf{V}_0^{-1}\mathbf{V}_1$. Let a matrix norm of $\mathbf{E}$, say $\rho(\mathbf{E}) = \|\mathbf{E}\|_\infty = \max|e_{ij}| = \max\frac{|c_{ij}|}{c_{ii}},\ i \neq j$, satisfies the inequality $\rho(\mathbf{E}) < 1$. Then the inverse matrix $\mathbf{V}^{-1} = (\mathbf{I} + \mathbf{E})^{-1}\mathbf{V}_0^{-1}$ can be expanded into the series (Horn and Johnson, 1986)

$$\mathbf{V}^{-1} = \sum_{m=0}^{\infty}(-\mathbf{E})^m \mathbf{V}_0^{-1}. \quad (12)$$

After restricting the series by the second term, we obtain

$$\mathbf{V}^{-1} \approx (\mathbf{I} - \mathbf{E})\mathbf{V}_0^{-1} = \left(\mathbf{I} - \mathbf{V}_0^{-1}\mathbf{V}_1\right)\mathbf{V}_0^{-1}. \quad (13)$$

Substituting this expression into equation (3) yields

Zero- and first-order approximations for least squares estimation of seismic signal

$$s = c^{-1}\mathbf{f}^H\mathbf{D}^{-1}\left(\begin{array}{l}\mathbf{I} - \sum_{l=1}^{L}\dfrac{\mathbf{g}_l\mathbf{g}_l^H\mathbf{D}^{-1}}{c_{ll}} \\ + \sum_{l=1}^{L}\mathbf{g}_l\sum_{k\neq l}\dfrac{\mathbf{g}_l^H\mathbf{D}^{-1}\mathbf{g}_k\mathbf{g}_k^H\mathbf{D}^{-1}}{c_{ll}c_{kk}}\end{array}\right)\mathbf{u}\,, \tag{14}$$

where

$$c = c_s - \sum_{l=1}^{L}\frac{|c_{sl}|^2}{c_{ll}} + \sum_{l=1}^{L}\sum_{k\neq l}\frac{c_{sl}c_{lk}c_{sk}^*}{c_{ll}c_{kk}}\,, \tag{15}$$

$$c_s = \mathbf{f}^H\mathbf{D}^{-1}\mathbf{f} = \sum_{i=1}^{N}\frac{a_i^2}{\sigma_i^2}\,, \tag{16}$$

$$c_{sl} = \mathbf{f}^H\mathbf{D}^{-1}\mathbf{g}_l = \sum_{k=1}^{N}\frac{a_k b_{kl}}{\sigma_k^2}\exp\left[i\omega\left(\tau_{(s)k} - \tau_{(r)kl}\right)\right]. \tag{17}$$

Following equation (14), we should first estimate and then subtract from the data **u** all the coherent noise wavetrains. For this aim, the shape of each coherent noise is estimated using optimum weighted stacking $c_{ll}^{-1}\mathbf{g}_l^H\mathbf{D}^{-1}\mathbf{u}$ (Tyapkin and Ursin, 2005) performed with reference to the variance of random noise and the amplitudes and arrival times of this coherent noise. Then it is multiplied by $\mathbf{g}_l$ in order to obtain the ultimate estimate of the *l*th coherent noise, with its amplitudes and arrival times on all the traces. The third term from the parentheses on the right-hand side of equation (14) is intended to compensate for the entire distortion caused by the mutual impact of all the coherent noise wavetrains. Indeed, when estimating the *l*th coherent noise, all the other $(k \neq l)$ wavetrains take part. In this process, the *k*th wavetrain, the estimate of which is $c_{kk}^{-1}\mathbf{g}_k\mathbf{g}_k^H\mathbf{D}^{-1}\mathbf{u}$, participates in the estimation of the *l*th coherent noise in accordance with the formula $c_{ll}^{-1}c_{kk}^{-1}\mathbf{g}_l\mathbf{g}_l^H\mathbf{D}^{-1}\mathbf{g}_k\mathbf{g}_k^H\mathbf{D}^{-1}\mathbf{u}$. Note there is no need in such a compensator if the coherent noise is single (L = 1). After subtracting all the coherent noise wavetrains, the residual data undergoes optimum weighted stacking $c^{-1}\mathbf{f}^H\mathbf{D}^{-1}$ with regard to the variance of random noise and the amplitudes and arrival times of the signal. When coherent noise is absent, $c = c_s$ (Tyapkin and Ursin, 2005). In general case c in equation (15) accounts for the corruptive impact on the signal of both subtracting coherent noise and compensating for the mutual influence of all the coherent noise wavetrains. It is easy to show by substituting the signal $\mathbf{f}s$ instead of the entire data **u** in equation (14).

We consider the procedure described by equation (14) *the first-order approximation* in contrast to *the zero-order approximation* when the mutual impact of all the coherent noise wavetrains is neglected (Tyapkin et al., 2007).

Synthetic data examples

In practice, it is difficult to find a reliable source of information on the variances of random noise and the amplitudes of coherent components of the seismic data. For this reason, when testing the method, we used synthetic data with these parameters being trace-independent. In this case optimum weighted stack turns into straight stack.

The proposed method is demonstrated on a synthetic data set containing a signal aligned in time. The signal is contaminated with two linear coherent noise wavetrains of dips 1ms and -1ms per trace and of various relative, in regard to the signal, amplitudes ranging from 1 to 16. The waveforms of all the coherent components were generated by convolution of independent stochastic processes with a 20Hz Ricker wavelet. Each synthetic record consists of 21 traces.

Figure 1 shows the entire record along with its components for the noise-to-signal ratio 4:1. The amplification factor for visualization of the coherent noise and entire data was chosen as one fourth of that for the signal.

In Fig. 2a, we demonstrate the result of subtracting coherent noise from the data depicted in Fig. 1d using the zero-order approximation. It is clearly seen that the residual record differs considerably from the ‘pure’ signal, ten traces of which are put in the right-hand side of each record here and in Fig. 3 for comparison. This distinction is also confirmed by the difference between the residual data and the signal (Fig. 2b), which has higher amplitudes than the signal. For this reason, the ultimate signal approximation yields a poor result and one can see a contrasting boundary between the signal and its estimate in Fig. 2c.

A much better result is obtained using the first-order approximation (Fig. 3). This advanced procedure subtracts coherent noise almost perfectly (Fig. 3a), which is also confirmed by much lower amplitudes of the difference between the result of noise subtraction and the signal (Fig. 3b). As a result, one can hardly find the boundary between the signal and its estimate in Fig. 3c.

The advantage of the first-order approximation over the zero-order approximation is demonstrated quantitatively with the coefficient of correlation between the signal and its estimates as a function of the noise-to-signal ratio (Fig. 4). It is always higher for the advanced procedure.

Conclusions

The method developed can be a valuable tool in processing surface and borehole seismic data. The results of its testing on synthetic data sets, some of which are demonstrated, indicate that in many circumstances when estimating signal contaminated by severe coherent noise our method can significantly outperform simplified approaches and may therefore be prescribed as a better choice than conventional processes. The aim of our future work is to supply the proposed method with reliable estimates of the required parameters, without which it can not operate effectively.

Zero- and first-order approximations for least squares estimation of seismic signal

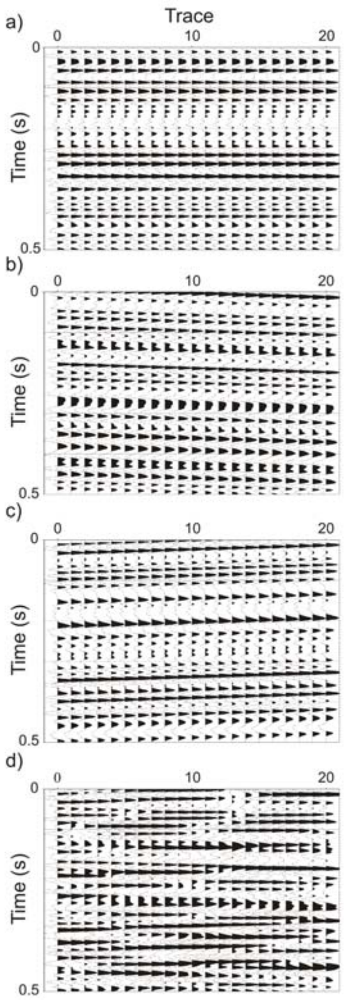

Figure 1: (a) Signal, (b) positively and (c) negatively dipping coherent noise and (d) entire data for the noise-to-signal ratio 4:1

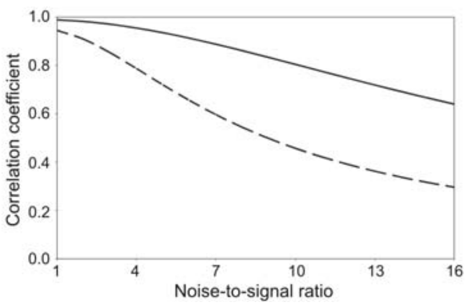

Figure 4: Correlation coefficients of the signal and its estimates obtained using advanced (solid line) and simplified (dashed line) procedures as a function of the noise-to-signal ratio

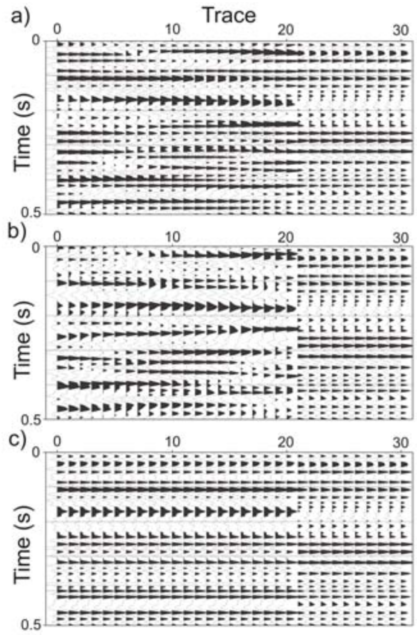

Figure 2: (a) Result of subtracting the coherent noise using the zero-order approximation, (b) difference between the residual data and the signal, (c) final signal estimate

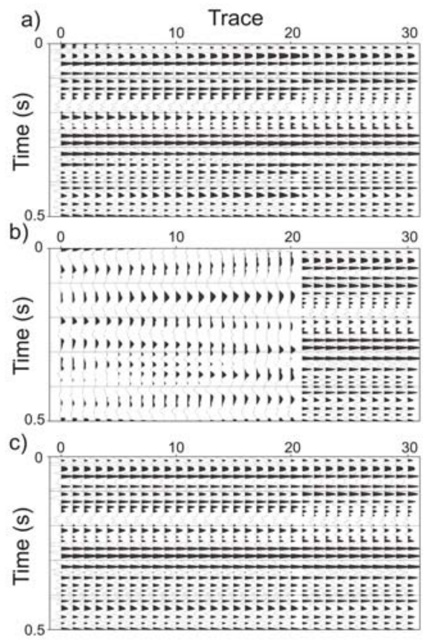

Figure 3: (a) Result of subtracting the coherent noise using the first-order approximation, (b) difference between the residual data and the signal, (c) final signal estimate

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2009 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Bekara, M., and M. Van der Baan, 2006, Local SVD/ICA for signal enhancement of pre-stack seismic data: 68th Annual Conference and Exhibition, EAGE, Extended Abstracts, D038.

Cassano, E., and F. Rocca, 1974, Multichannel filters without mixing effects: Geophysical Prospecting, **22**, 330–344.

Gounon, P. Y., J. I. Mars, and D. Goncalves, 1998, Wideband spectral matrix filtering: 68th Annual International Meeting, SEG, Expanded Abstracts, 1150–1153.

Hanna, M. T., and M. Simaan, 1985, Optimum simultaneous suppression of coherent and random noise in seismic data: 55th Annual International Meeting, SEG, Expanded Abstracts, 627–629.

Helstrom, C. W., 1968, Statistical theory of signal detection: Pergamon Press, Inc.

Horn, R. A., and C. R. Johnson, 1986, Matrix analysis: Cambridge Univ. Press.

March, D. W., and A. D. Bailey, 1983, A review of the two-dimensional transform and its use in seismic processing: First Break, **1**, 9–21.

Meyerhoff, H. J., 1966, Horizontal stacking and multichannel filtering applied to common depth point seismic data: Geophysical Prospecting, **14**, 441–454.

Mitchell, A. R., and P. G. Kelamis, 1990, Efficient tau-*p* hyperbolic velocity filtering: Geophysics, **55**, 619–625.

Schoenberger, M., 1996, Optimum weighted stack for multiple suppression: Geophysics, **61**, 891–901.

Trickett, S., 2003, *F-xy* eigenimage noise suppression: Geophysics, **68**, 751–759.

Tyapkin, Y., and B. Ursin, 2005, Optimum stacking of seismic record with irregular noise: Journal of Geophysics and Engineering, **2**, 177–187.

Tyapkin, Y., B. Ursin, Y. Roganov, I. Nekrasov, and G. Shumlyanskaya, 2007, Least-squares signal estimation with complicated mathematical models of seismic data: 69th Annual Conference and Exhibition, EAGE, Extended Abstracts, P026.

Ulrych, T. J., M. D. Sacchi, and J. M. Graul, 1999, Signal and noise separation: Art and science: Geophysics, **64**, 1648–1672.

Vrabie, V. D., J. I. Mars, and J. L. Lacoume, 2004, Modified singular value decomposition by means of independent component analysis: Signal Processing, **84**, 645–652.

Zhu, W., P. G. Kelamis, and Q. Liu, 2004, Acquisition/processing—Linear noise attenuation using local radial trace median filtering: The Leading Edge, **23**, 728–737.