Adaptive subtraction of multiples using the L_1-norm

A. Guitton[1] and D. J. Verschuur[2]*

[1]*Stanford Exploration Project, Mitchell Building, Department of Geophysics, Stanford University, Stanford, CA 94305-2215, USA, and*
[2]*Delft University of Technology, Laboratory of Acoustical Imaging and Sound Control, PO Box 5046, 2600 GA Delft, The Netherlands*

ABSTRACT

A strategy for multiple removal consists of estimating a model of the multiples and then adaptively subtracting this model from the data by estimating shaping filters. A possible and efficient way of computing these filters is by minimizing the difference or misfit between the input data and the filtered multiples in a least-squares sense. Therefore, the signal is assumed to have minimum energy and to be orthogonal to the noise. Some problems arise when these conditions are not met. For instance, for strong primaries with weak multiples, we might fit the multiple model to the signal (primaries) and not to the noise (multiples). Consequently, when the signal does not exhibit minimum energy, we propose using the L_1-norm, as opposed to the L_2-norm, for the filter estimation step. This choice comes from the well-known fact that the L_1-norm is robust to 'large' amplitude differences when measuring data misfit. The L_1-norm is approximated by a hybrid L_1/L_2-norm minimized with an iteratively reweighted least-squares (IRLS) method. The hybrid norm is obtained by applying a simple weight to the data residual. This technique is an excellent approximation to the L_1-norm. We illustrate our method with synthetic and field data where internal multiples are attenuated. We show that the L_1-norm leads to much improved attenuation of the multiples when the minimum energy assumption is violated. In particular, the multiple model is fitted to the multiples in the data only, while preserving the primaries.

INTRODUCTION

A classical approach to attenuating multiples consists of building a multiple model (see e.g. Verschuur, Berkhout and Wapenaar 1992; Berkhout and Verschuur 1997), and adaptively subtracting this model from the data, which is contaminated with multiples, by estimating shaping filters (Dragoset 1995; Liu, Sen and Stoffa 2000; Rickett, Guitton and Gratwick 2001). The estimation of the shaping filters is usually carried out in a least-squares sense making these filters relatively easy to compute. By using the L_2-norm, we implicitly assume that the resulting signal, after the filter estimation step, is orthogonal to the noise and has minimum energy. These assumptions might not hold, and other methods, such as pattern-based approaches (Spitz 1999; Guitton *et al.* 2001), have been proposed to avoid these limitations. For instance,

when a strong primary is surrounded by weaker multiples, the multiple model will match both the noise (multiples) and the signal (primaries), such that the difference between the data and the filtered multiple model is a minimum in a least-squares sense. Consequently, some primary energy might leak into the estimated multiples and *vice versa*. We therefore need to find a new criterion or norm for the filter estimation step.

We propose estimating the shaping filters with the L_1-norm instead of the L_2-norm, thus removing the necessity for the signal to have minimum energy. This choice is driven by the simple fact that the L_1-norm is robust to 'outliers' (Claerbout and Muir 1973) and large amplitude anomalies. Because the L_1-norm is singular where any residual component vanishes, we use a hybrid L_1/L_2-norm that we minimize with an iteratively reweighted least-squares (IRLS) method. This method is known to give an excellent approximation of the L_1-norm (Gersztenkorn, Bednar and Lines 1986; Scales and Gersztenkorn 1987; Bube and Langan 1997; Zhang,

---

*E-mail: d.j.verschuur@ctg.tudelft.nl

Chunduru and Jervis 2000). The main property of the hybrid norm is that it is continuous and differentiable everywhere, while being robust for large residuals.

In the first section, we illustrate the limitations of the least-squares criterion using a simple 1D problem. We then introduce our proposed approach, based on the L_1-norm, to improving the multiple attenuation results. In a second synthetic example, we attenuate internal multiples with the L_2- and L_1-norms. Finally, we apply shaping filters to a multiple-contaminated gather from a seismic survey, showing that the L_1-norm leads to substantial attenuation of the multiples.

PRINCIPLES OF L_1-NORM AND L_2-NORM SUBTRACTION

In this section, we demonstrate with a 1D example that the attenuation of multiples with least-squares adaptive filtering is not effective when strong primaries are located in the neighbourhood. This simple example leads to a better understanding of the behaviour of our adaptive scheme in more complicated cases.

Shaping filters and the L_2-norm

In Fig. 1, a simple 1D problem is considered. Figure 1(a) shows four events corresponding to one primary (on the left) and three multiples (on the right). Note that the primary has a larger amplitude than the multiples. Figure 1(b) shows a mul-

tiple model that corresponds exactly to the real multiples. For L_2-norm subtraction the goal is to estimate a shaping filter $\mathbf{f}$ that minimizes the objective function,

$$e_2(\mathbf{f}) = \|\mathbf{d} - \mathbf{Mf}\|_2^2, \tag{1}$$

where $\mathbf{M}$ is the matrix representing the convolution with the time series for the multiple model (Fig. 1b) and $\mathbf{d}$ is the time series for the data (Fig. 1a).

If we estimate the filter $\mathbf{f}$ with enough degrees of freedom (enough coefficients) to minimize (1), we obtain the estimated primaries, i.e. $\mathbf{d} - \mathbf{Mf}$ (Fig. 2a), and the estimated multiples, i.e. $\mathbf{Mf}$ (Fig. 2b). The estimated primary signal does not resemble the primary in Fig. 1(a). In Fig. 3, the corresponding shaping filter is shown. Note that this filter is not a unit spike at $lag = 0$ as expected. The problem stems from the least-squares criterion which yields an estimated signal that, by definition, has minimum energy. In this 1D case, the total energy in the estimated signal (Fig. 2a) is $e_2 = 2.4$, which is less than the total energy of the primary alone ($e_2 = 4$). This is the fundamental problem if we use the L_2-norm to estimate the shaping filter. In the next section, we show that it is better to use the L_1-norm if the multiples and the primaries are not orthogonal in the L_2-norm sense.

Shaping filters and the L_1-norm

The strong primary in Fig. 1 can be seen as an outlier that receives much attention during the L_2 filter estimation.
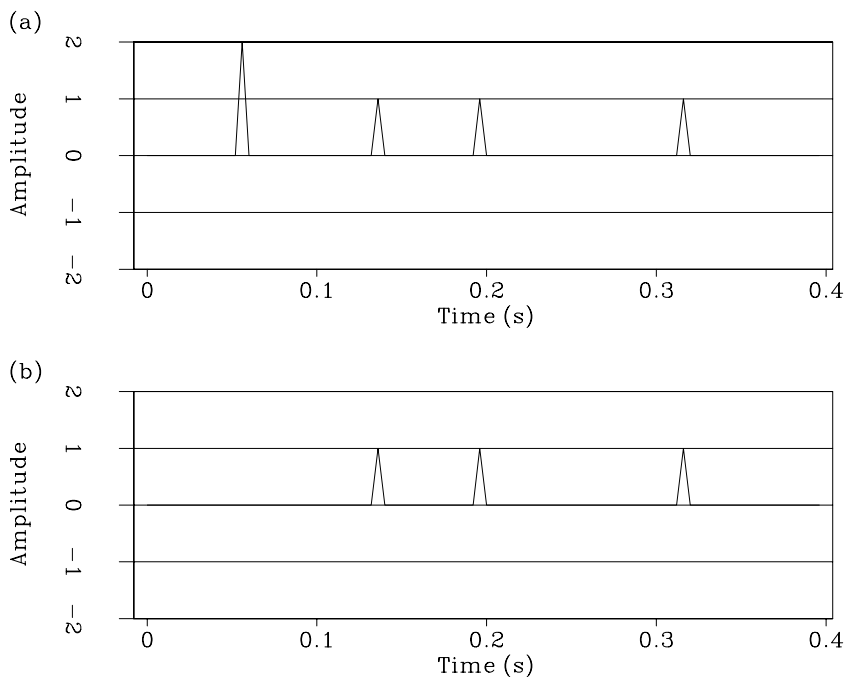

Figure 1 (a) The data with one primary at 0.06 s and three multiples at 0.14 s, 0.2 s and 0.32 s. (b) The multiple model that we want to adaptively subtract from (a).

Figure 2 (a) The signal estimated with the L₂-norm. (b) The noise estimated with the L₂-norm.
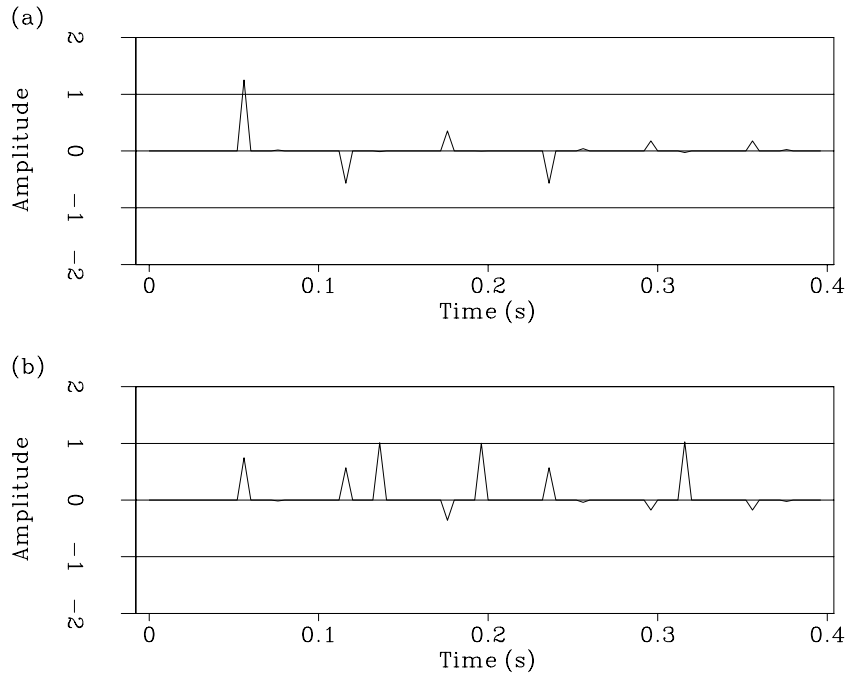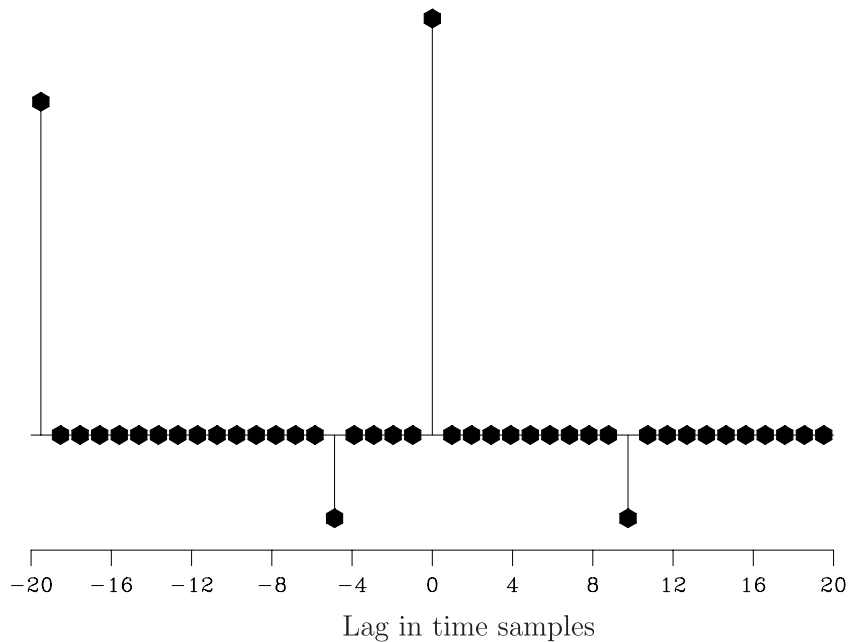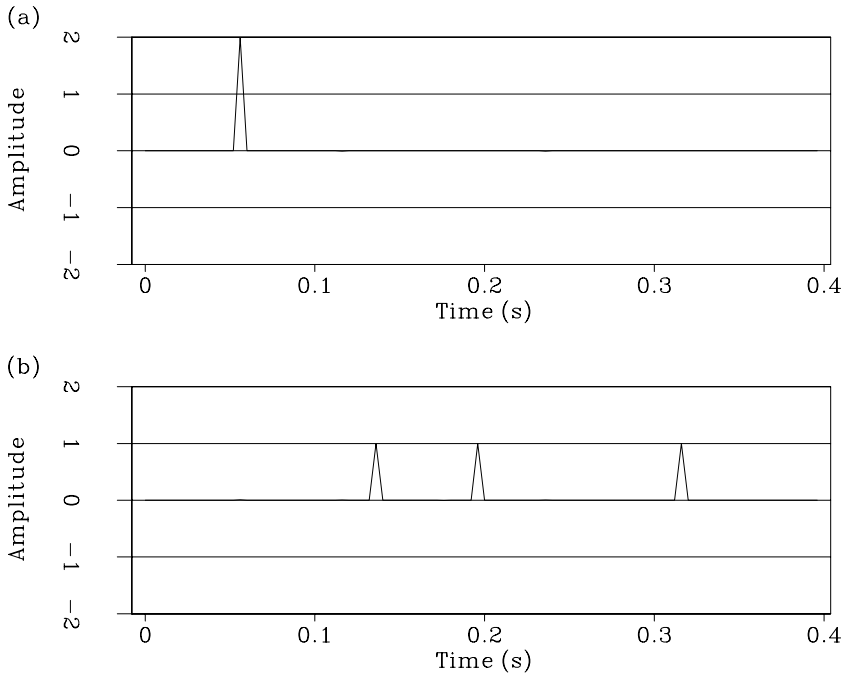

Figure 3 Shaping filter estimated for the 1D problem with the L₂-norm. One lag is equivalent to one time sample (0.004 s). This filter is not a single spike at *lag* = 0. The maximum value of the filter is one at zero lag and the minimum value is −0.2 at lags −5 and +9.

Consequently, some of the signal we want to preserve leaks into the noise. Because the L_1-norm is robust to outliers, we propose using it to estimate the filter coefficients. This insensitivity to a large amount of 'noise' has a statistical interpretation: robust measures are related to long-tailed density functions in the same way that L_2 is related to the short-tailed Gaussian density function (Tarantola 1987). In this section,

we show that the L_1-norm solves the problem referred to in the preceding section.

Our goal now is to estimate a shaping filter **f** that minimizes the objective function,

$$e_1(\mathbf{f}) = |\mathbf{d} - \mathbf{M}\mathbf{f}|_1. \tag{2}$$

(a)

(b)

Figure 4 (a) The signal estimated with the L_1-norm. (b) The noise estimated with the L_1-norm.
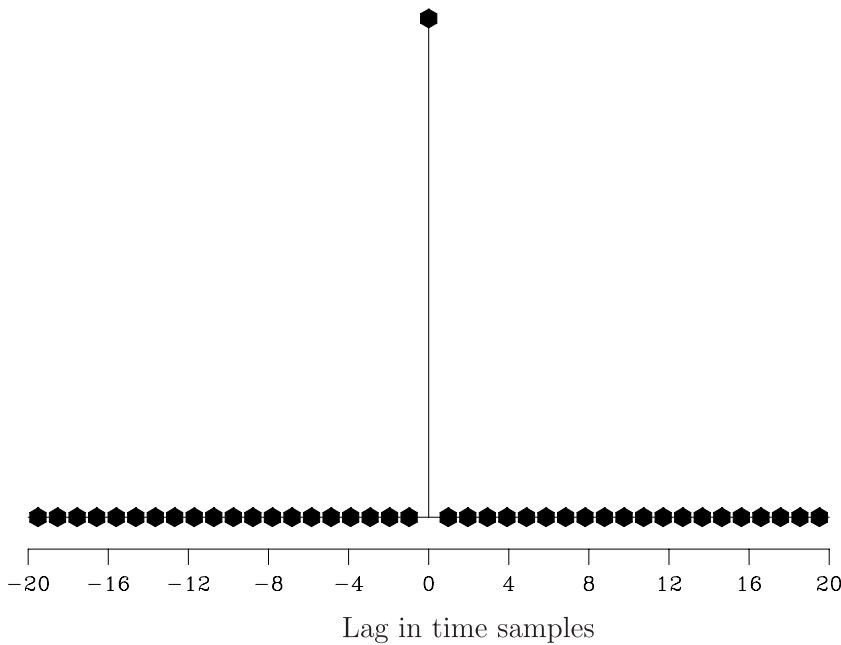

Lag in time samples

Figure 5 Shaping filter estimated for the 1D problem with the L_1-norm. One lag is equivalent to one time sample (0.004 s). This filter is a single spike at *lag* = 0 with amplitude value 1.0.

The function in (2) is singular where any residual component vanishes, implying that the derivative of $e_1(\mathbf{f})$ is not continuous everywhere. Unfortunately, most of our optimization techniques, e.g. conjugate-gradient or Newton methods, assume that the first derivative of the objective function is continuous in order to find its minimum. Therefore, specific techniques have been developed either to minimize or to approximate the L_1-norm. For instance, various approaches based on linear programming have been used with success (see e.g. Barrodale and Roberts 1980). Other robust measures, such as the Huber norm (Huber 1973), can also be considered with an appropriate minimization scheme (Guitton and Symes 1999).

Alternatively, our implementation is based on the minimization of a hybrid L_1/L_2-norm with an iteratively reweighted least-squares (IRLS) method (Gersztenkorn *et al.* 1986; Scales
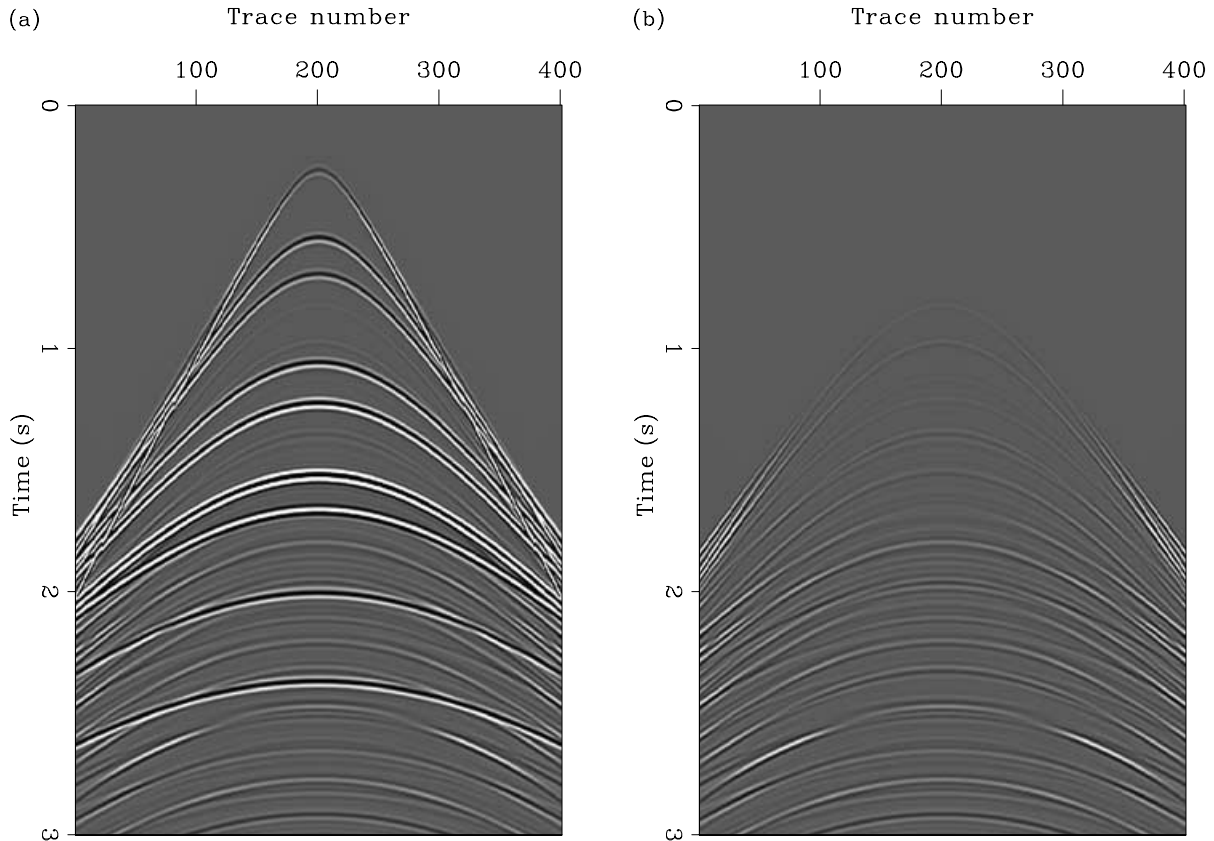
(a)

(b)

Figure 6 (a) A synthetic shot gather containing many internal multiples. (b) The internal multiple model, which exactly matches the internal multiples in (a).
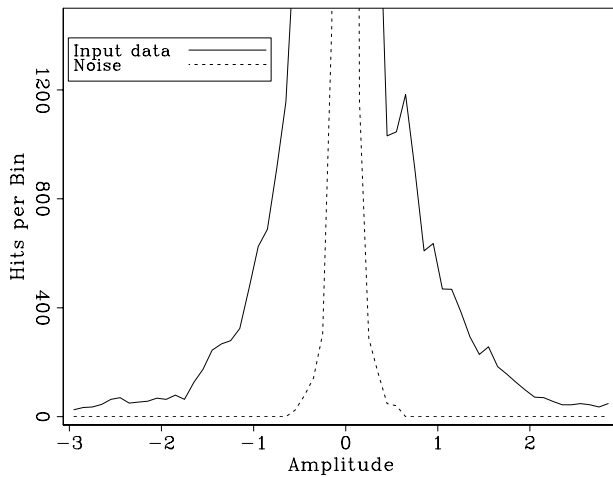

Figure 7 Histograms of the input data (Fig. 6a) and of the noise (Fig. 6b). The multiples have a much weaker amplitude distribution and the L_1-norm should be used.

and Gersztenkorn 1988; Scales *et al.* 1988; Bube and Langan 1997). This technique is known to give a good approximation of the L_1-norm. In this case, the objective function we minimize becomes

$$e_1(\mathbf{f}) = \|\mathbf{W}(\mathbf{d} - \mathbf{M}\mathbf{f})\|_2^2, \tag{3}$$

with

$$\mathbf{W} = \mathbf{diag}\left(\frac{1}{\left(1 + r_i^2/\epsilon^2\right)^{1/4}}\right), \tag{4}$$

where $r_i = \mathbf{d} - \mathbf{M}\mathbf{f}_i$ is the residual for one component of the data space, and ϵ is a constant chosen *a priori*. The significance of ϵ and how it is chosen is described later. With this particular choice of $\mathbf{W}$, minimizing e_1 is equivalent to minimizing

$$Q(\mathbf{f}) = \sum_{i=1}^{N} q(r_i) = \sum_{i=1}^{N} \left(\sqrt{1 + (r_i/\epsilon)^2} - 1\right), \tag{5}$$

where N is the number of data points (see Bube and Langan 1997). For any given residual r_i, we have

$$q(r_i) \approx \begin{cases} \frac{1}{2}(r_i/\epsilon)^2, & \text{for } |r_i|/\epsilon \text{ small,} \\ |r_i|/\epsilon, & \text{for } |r_i|/\epsilon \text{ large.} \end{cases} \tag{6}$$
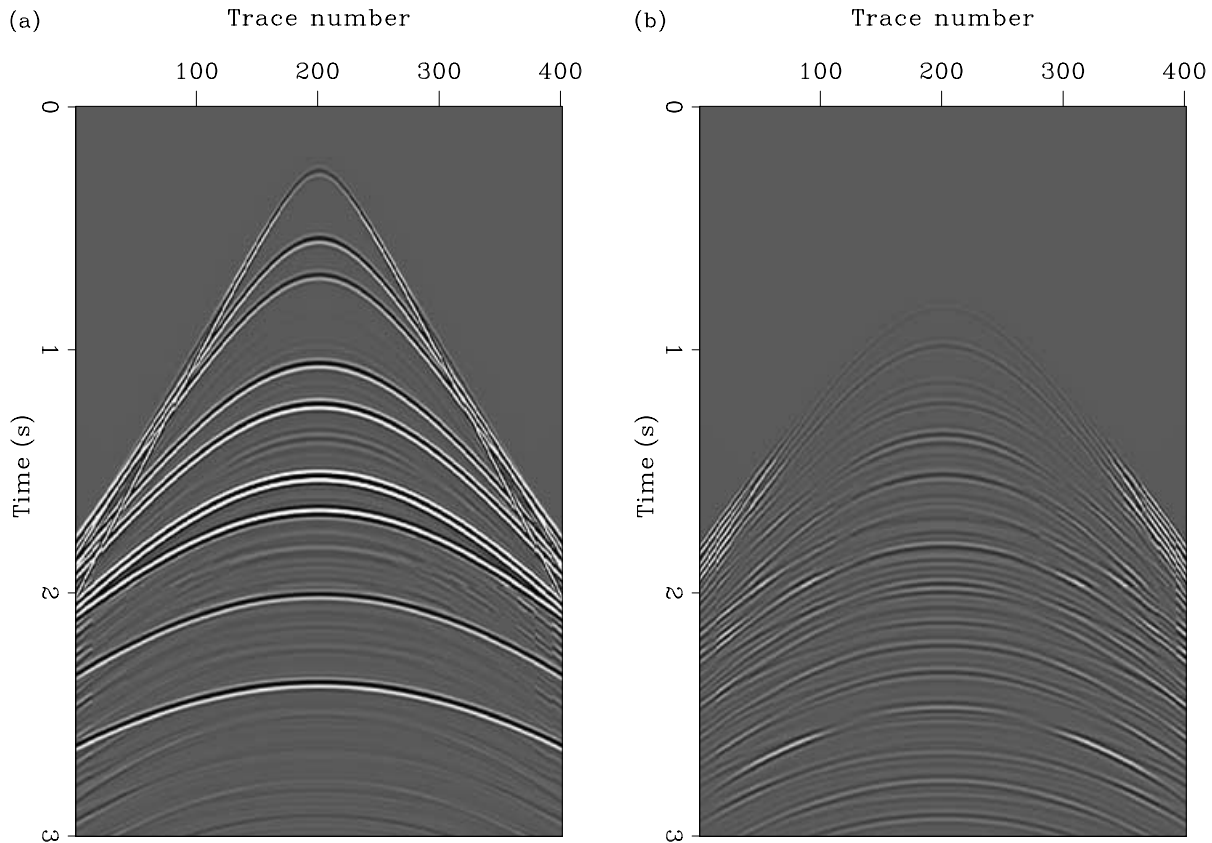
Figure 8 (a) The estimated primaries with the L_2-norm. (b) The estimated internal multiples with the L_2-norm. Ideally, (b) should look like Fig. 6(b), but in this case it does not.

Hence, we obtain L_1 treatment of large residuals and L_2 treatment of small residuals with a smooth transition between the two norms. This transition is defined by ϵ. Note that the objective function in (3) is non-linear because the weighting matrix $\mathbf{W}$ is a function of the residual for every data point. Therefore, we use a non-linear technique to minimize (3), i.e. IRLS. This method solves the non-linear problem with piecewise linear steps. For each linear step, $\mathbf{W}$ is kept constant.

We now describe the implementation of IRLS in more detail. We update the weighting operator $\mathbf{W}$ every five iterations. Within each linear step a conjugate-gradient (CG) solver is used, which makes the minimization of the hybrid norm very fast. The CG solver is reset for each new weighting function, so that the first iteration of each new least-squares problem is a steepest-descent step. In addition, the last solution $\mathbf{f}$ of the previous linear problem is used as an initial guess $\mathbf{f}_0$ for the next five iterations. Note that the total cost per iteration of the L_1 method is equal to the cost of the L_2 method because both use a CG solver. Finally, because we solve a non-linear

problem with IRLS, the L_1-norm requires twice as many iterations as the L_2-norm. This cost increment is quite reasonable and makes the L_1-norm affordable.

One important parameter in the definition of the weighting function $\mathbf{W}$ is ϵ. In (6), this parameter controls the transition between the L_1-norm and the L_2-norm (Bube and Langan 1997). Bube and Langan (1997) proposed computing ϵ as a function of the standard deviation of the residual $\mathbf{d} - \mathbf{Mf}$. They proposed that if the standard deviation of the residual varies for different points, then there should be one ϵ per residual component r_i. Although mathematically supported, in our opinion this is not very practical. Therefore, we calculate one ϵ only as follows:

$$\epsilon = \frac{\max|\mathbf{d}|}{100}. \tag{7}$$

This choice has been proved efficient by a few authors (see e.g. Darche 1989; Nichols 1994). Note that a comparison of our choice of ϵ with that proposed by Bube and Langan
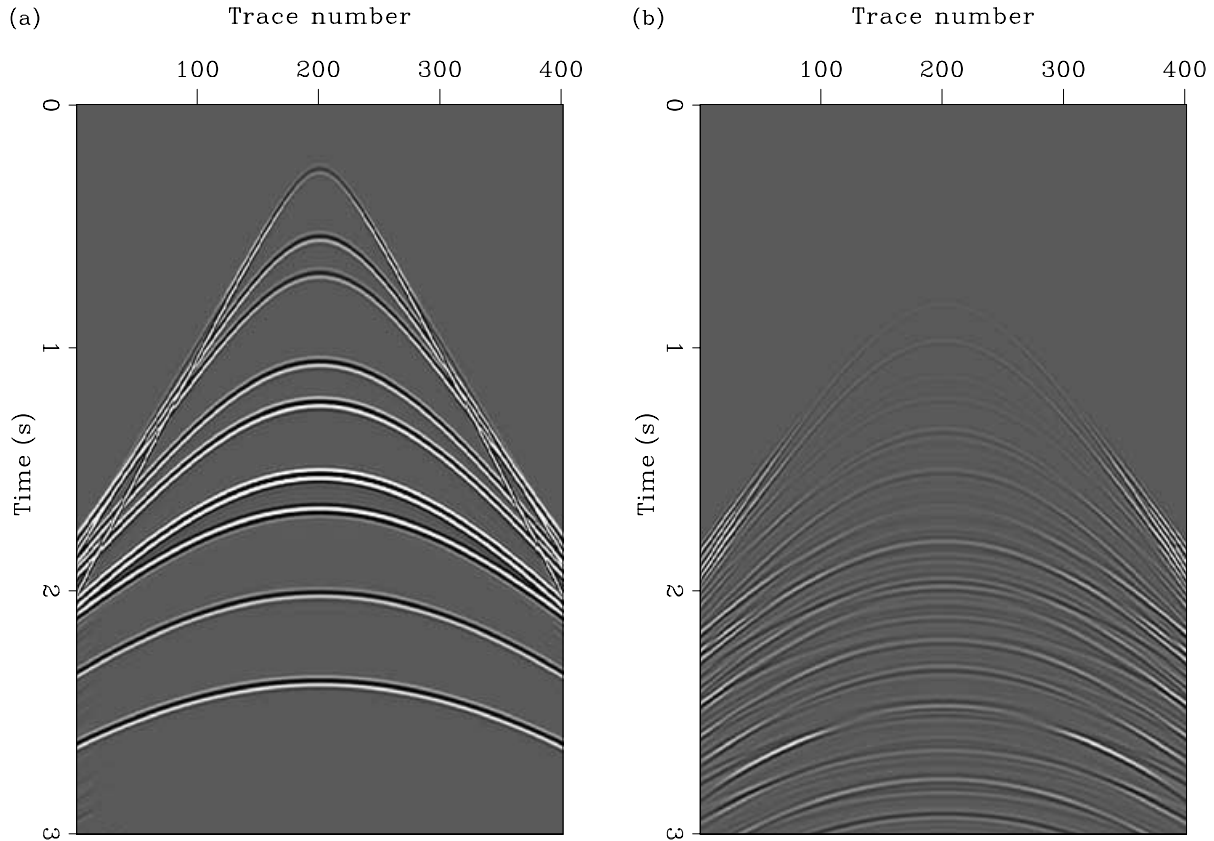
(a)

(b)

Figure 9 (a) The estimated primaries with the L_1-norm. (b) The estimated internal multiples with the L_1-norm. Apart from some edge effects, (b) closely resembles Fig. 6(b). The adaptive subtraction was very successful.

(1997) could be interesting, but goes beyond the scope of this paper.

From now on, we refer to our hybrid L_1/L_2-norm, minimized with the IRLS method, as the L_1-norm. In the next section, we show that our implementation leads to the desired result for the filter estimation problem.

Results from a simple 1D example

For the simple 1D example under consideration, the filter coefficients are now estimated with the L_1-norm using the IRLS method. In Fig. 4, we show the result of the adaptive subtraction when the L_1-norm is used to estimate the shaping filter (equation (3) with small ϵ). The estimated signal in Fig. 4(a) resembles the true signal very well, as does the estimated noise. It is easy to check that the energy (in an L_1 sense) in Fig. 4(a) ($e_1 = 2$) is less than the energy (in an L_1 sense) in Fig. 2(a) ($e_1 = 3.2$). Figure 5 shows the shaping filter associated with the L_1-norm. This filter is a unit spike at $lag = 0$. This simple

1D example demonstrates that the L_1-norm should be utilized each time the estimated noise and the desired primaries are not orthogonal in the L_2 sense. In the following section, we show another synthetic example where internal multiples are attenuated.

2D DATA EXAMPLE: ATTENUATION OF INTERNAL MULTIPLES

In this section we illustrate the efficiency of the L_1-norm when internal multiples are attenuated in 2D.

The synthetic data

Figure 6(a) shows a synthetic shot gather for a 1D medium. This gather is corrupted with internal multiples only. Figure 6(b) shows the internal multiple model. In order to focus on the multiple subtraction only, and not on the prediction, this internal multiple model is exact and could be directly
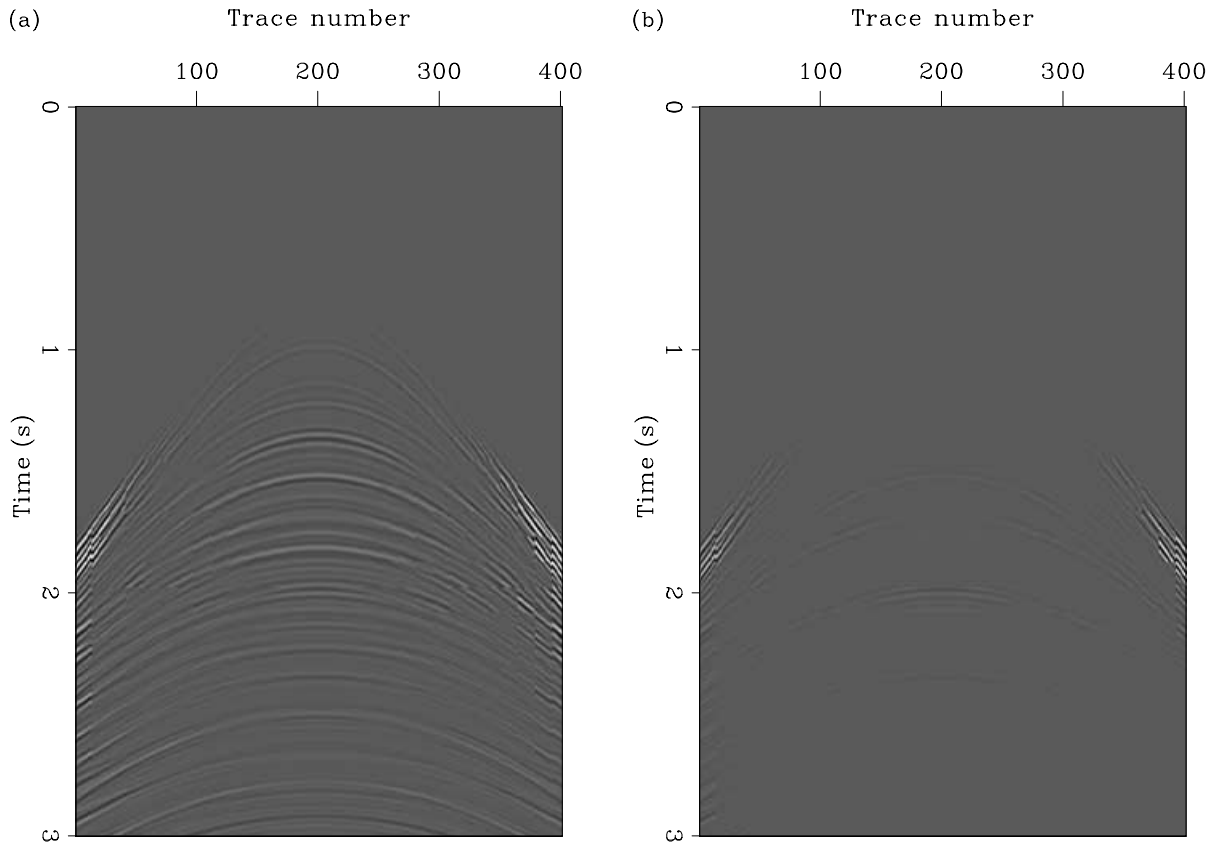
Figure 10 (a) The difference between the exact multiples (Fig. 6b) and the subtracted multiples with the L_2-norm. (b) The difference between the exact multiples (Fig. 6b) and the subtracted multiples with the L_1-norm. The L_1-norm gives a more accurate multiple attenuation result.

subtracted from the data in Fig. 6(a). Note that the amplitude of the internal multiples is significantly less than the amplitude of the primaries, making the L_2-norm unsuitable for estimating the shaping filters. Figure 7 shows the histograms of both the data and the internal multiples. The density function of the noise is much narrower than that of the data, indicating that the L_1-norm should be used.

Adaptive filtering with non-stationary helical filters

To handle the inherent non-stationarity of seismic data, we estimate a bank of non-stationary filters using helical boundary conditions (Mersereau and Dudgeon 1974; Claerbout 1998). This approach has been successfully used by Rickett *et al.* (2001) to attenuate surface-related multiples. As described above (equations (2)–(7)), we use IRLS to approximate the L_1-norm and a standard conjugate-gradient solver with the L_2-norm. The filter coefficients vary smoothly across the output space, thanks to preconditioning of the problem (Crawley

2000; Rickett *et al.* 2001). In the following results, the non-stationary filters are 1D. We estimate the same number of coefficients per filter with the L_2- and L_1-norms.

Adaptive subtraction results

Figure 8(a) shows the estimated primaries when the L_2-norm is used to compute the shaping filters. Figure 8(b) shows the estimated internal multiples. As expected, because of the non-orthogonality of the signal (primaries) and the noise (multiples) in the L_2 sense, the adaptive subtraction fails and we retrieve the behaviour explained in the preceding section with the 1D example: due to predicted internal multiples being close to relatively strong primaries, part of the energy of these multiples is matched with the primaries, resulting in a non-optimum subtraction, so that the subtraction result has less energy in the L_2 sense than the desired output. Next, in Fig. 9, we see the beneficial effects of the L_1-norm. Figure 9(a) shows the estimated primaries and Fig. 9(b) shows the estimated
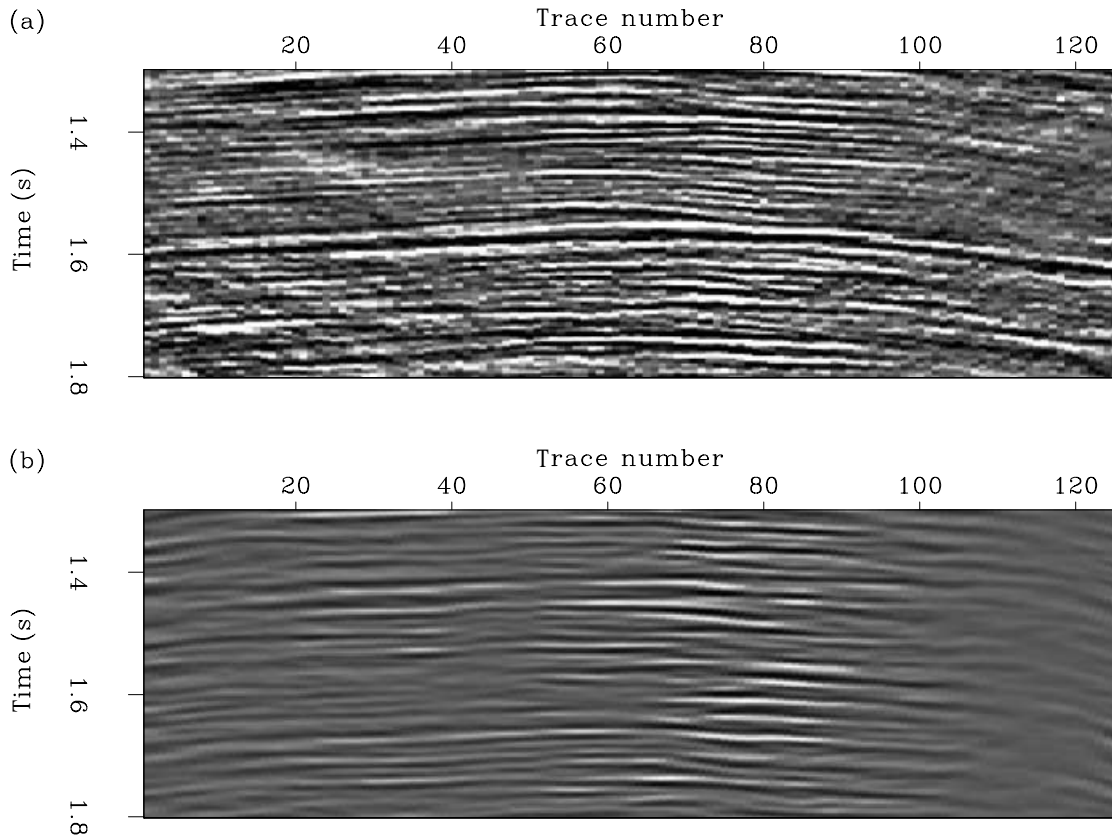
(a)

(b)

Figure 11 (a) Stack of land data contaminated with multiples. (b) The predicted multiples computed with the data-driven modelling approach. The subtraction is performed post-stack.

multiples. The noise subtracted matches the internal multiple model in Fig. 6(b) very well, as anticipated. As a final comparison, Fig. 10 shows the difference between the internal multiple model (Fig. 6b) and the subtracted multiples with the two norms. The L_1-norm (Fig. 10b) matches the multiple model much better than the L_2-norm (Fig. 10a).

POST-STACK LAND DATA MULTIPLE REMOVAL EXAMPLE

In this section, we attenuate surface-related multiples in the post-stack domain, using shaping filters that we estimate with the L_2- and L_1-norms. These filters are non-stationary. Figure 11(a) shows the data, which is contaminated with multiples. Figure 11(b) shows the multiple model computed using the data-driven modelling approach (Kelamis and Verschuur 2000). Note that for this gather, the amplitude differences between the primaries and the multiples are not very great. Our goal is to illustrate the use of the L_1-norm in a more general case when surface-related multiples are present in the

data. We specifically focus on the event at 1.6 s in Fig. 11(a). This is a primary event that we want to preserve during the subtraction.

Adaptive subtraction results

The amplitude of the primary at 1.6 s is well preserved with the L_1-norm in Fig. 12(a). However, the amplitude of this primary is attenuated with the L_2-norm as shown in Fig. 12(b). Figure 13 shows a comparison between the subtracted multiples with the L_1- (Fig. 13a) and the L_2-norm (Fig. 13b). We conclude that the L_2-norm tends to subtract too much energy.

This last example proves that the estimation of shaping filters can always be carried out with the L_1-norm. An advantage of our inversion scheme and the objective function in (3) is that only one parameter (ϵ) controls the L_1–L_2 behaviour. Thus we can decide to switch from one norm to another very easily. Figure 14 shows a histogram of the input data and of the estimated noise with the L_1- and L_2-norms. The theory predicts that the distribution of the L_2 result should be Gaussian and
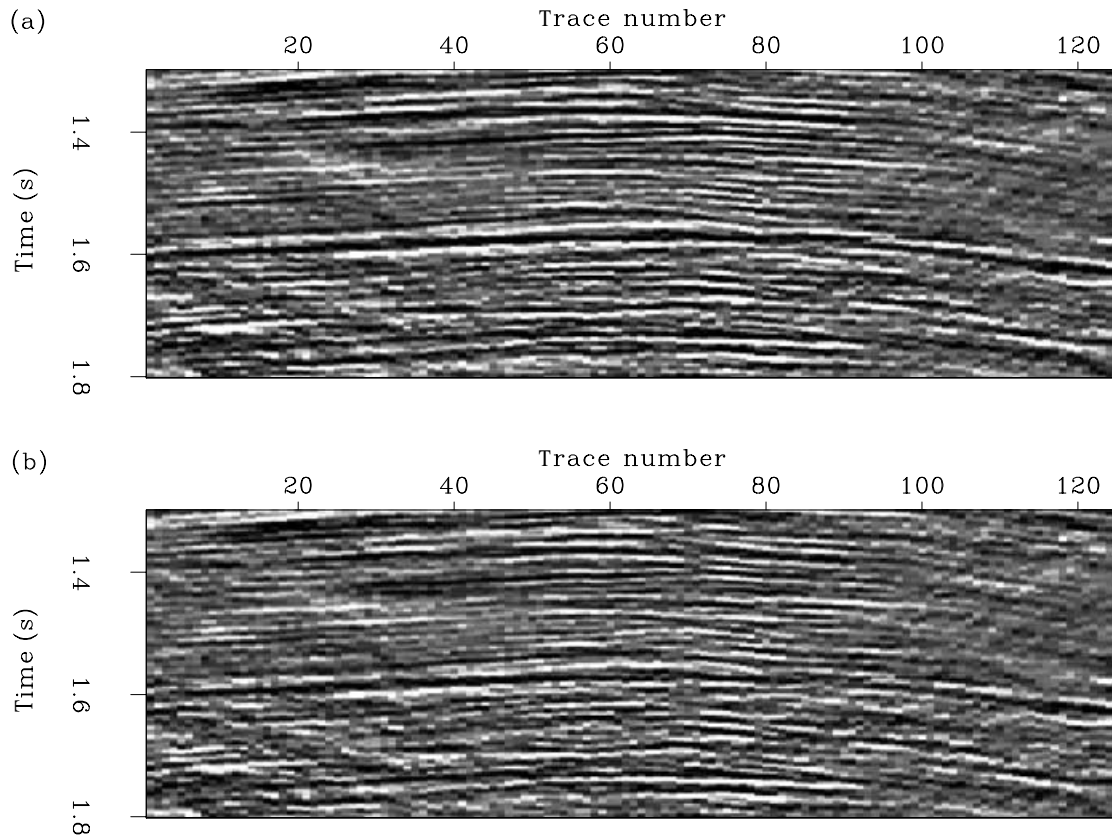
Figure 12 (a) The estimated primaries with the L_1-norm adaptive subtraction. (b) The estimated primaries with the L_2-norm subtraction. The primary at 1.6 s is very attenuated with the L_2-norm. The L_1 technique preserves its amplitude better.

that the distribution of the L_1 result should be exponential. Figure 14 corroborates this. Of course, we cannot discover directly whether the estimated multiples obtained with the L_1-norm are more similar to the actual multiples than those obtained with the L_2-norm. Our judgment is based only on qualitative considerations for a few known primary reflectors that we want to preserve.

Some efficiency considerations

As described above, the hybrid L_1–L_2 algorithm can be used to carry out both an L_2-norm subtraction (by choosing the parameter ϵ to be large) or a robust L_1-norm subtraction (by choosing ϵ to be small). In the algorithm used, the filters are allowed to change continuously along the time and space axes, and the data within a certain window around the output point will influence each filter. As stated before, the IRLS method used becomes more efficient in the case of L_2-norm subtraction (twice as fast) but is still 5–10 times slower than a conventional

linear least-squares algorithm using the Levinson scheme (see e.g. Robinson and Treitel 1980). However, the latter algorithm cannot handle smoothly varying, non-stationary filters, but requires an implementation with stationary filters in overlapping time-offset windows (as used by Verschuur and Berkhout 1997). It can lead to transition problems in the window overlap zones, resulting in non-optimum multiple suppression.

CONCLUSIONS

When a model of the multiples is adaptively subtracted from the data in a least-squares sense, we implicitly assume that the signal (primaries) has minimum energy and is orthogonal to the noise (multiples). This paper demonstrates that the minimum energy assumption might not hold and that another norm, the L_1-norm, should be used instead. The L_1-norm is approximated with a hybrid L_1/L_2-norm, which is minimized using an iteratively reweighted least-squares method. This method involves solving a non-linear problem with a
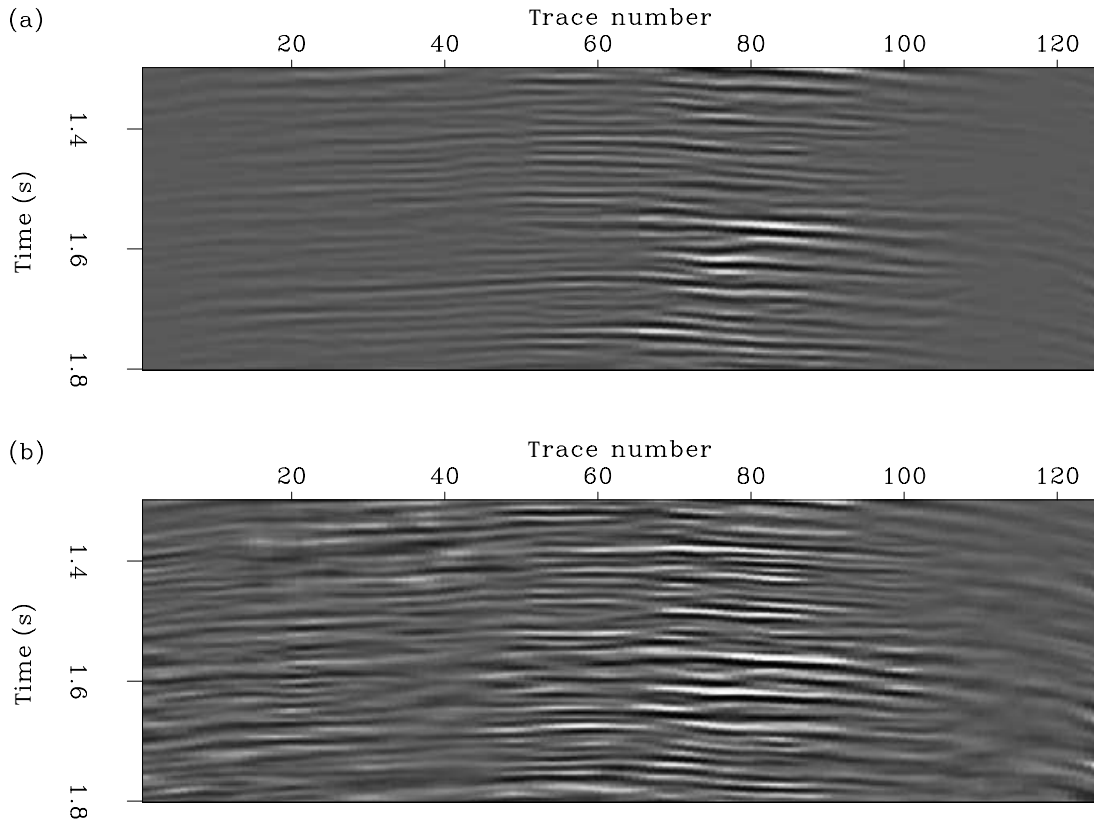
(a)

(b)

Figure 13 (a) The estimated multiples with the L_1-norm subtraction. (b) The estimated multiples with the L_2-norm subtraction. The L_2-norm tends to over-fit some multiples, which leads to some leaking of primaries in the estimated noise.
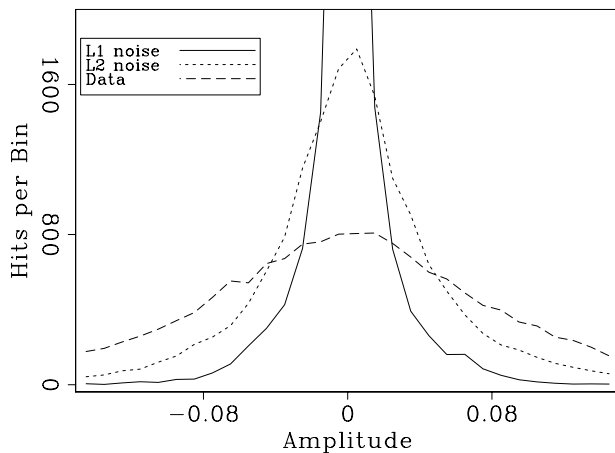

Figure 14 Histograms of the input data and of the estimated noise with the L_1- and L_2-norms. As predicted by the theory, the density function with the L_1-norm is much narrower than that with the L_2-norm.

series of linear steps. Both L_1- and L_2-norms are minimized with a conjugate-gradient solver but the L_1-norm requires twice as many iterations as the L_2-norm to achieve convergence. We demonstrated with 1D and 2D data examples that

our proposed scheme with the L_1-norm gives greatly improved multiple attenuation results when the desired primary signal does not exhibit minimum energy: the multiples are well suppressed and the primaries are preserved.

ACKNOWLEDGEMENTS

REFERENCES

Barrodale I. and Roberts F.D.K. 1980. Algorithm 552: Solution of the constrained l1 linear approximation problem. *ACM Transactions on Mathematical Software* **6**, 231–235.

Berkhout A.J. and Verschuur D.J. 1997. Estimation of multiple scattering by iterative inversion, part I: Theoretical considerations. *Geophysics* **62**, 1586–1595.

Bube K.P. and Langan R.T. 1997. Hybrid l1/l2 minimization with applications to tomography. *Geophysics* **62,** 1183–1195.

Claerbout J. 1998. Multidimensional recursive filters via a helix. *Geophysics* **63,** 1532–1541.

Claerbout J.F. and Muir F. 1973. Robust modeling with erratic data. *Geophysics* **38,** 820–844.

Crawley S. 2000. *Seismic trace interpolation with nonstationary prediction-error filters*. PhD thesis, Stanford University.

Darche G. 1989. Iterative L1 deconvolution. *SEP Annual Report* **61,** 281–301.

Dragoset W.H. 1995. Geophysical applications of adaptive-noise cancellation. 65th SEG meeting, Houston, USA, Expanded Abstracts, 1389–1392.

Gersztenkorn A., Bednar J.B. and Lines L.R. 1986. Robust iterative inversion for the one-dimensional acoustic wave equation. *Geophysics* **51,** 357–368.

Guitton A., Brown M., Rickett J. and Clapp R. 2001. Multiple attenuation using a t-x pattern-based subtraction method. 71st SEG meeting, San Antonio, Texas, USA, Expanded Abstracts, 1334–1337.

Guitton A. and Symes W.W. 1999. Robust and stable velocity analysis using the Huber function. 69th SEG meeting, Houston, USA, Expanded Abstracts, 1166–1169.

Huber P.J. 1973. Robust regression: Asymptotics, conjectures, and Monte Carlo. *Annals of Statistics* **1,** 799–821.

Kelamis P.G. and Verschuur D.J. 2000. Surface-related multiple elimination on land seismic data – strategies via case studies. *Geophysics* **63,** 719–734.

Liu F., Sen M.K. and Stoffa P.L. 2000. Dip selective 2-D multiple attenuation in the plane-wave domain. *Geophysics* **65,** 264–274.

Mersereau R.M. and Dudgeon D.E. 1974. The representation of two-dimensional sequences as one-dimensional sequences. *IEEE Transactions on Acoustics, Speech and Signal Processing* **22,** 320–325.

Nichols D. 1994. Velocity-stack inversion using Lp norms. *SEP Report* **82,** 1–16.

Rickett J., Guitton A. and Gratwick D. 2001. Adaptive multiple suppression with preconditioned non-stationary shaping filters. 63rd EAGE Conference, Amsterdam, The Netherlands, Extended Abstracts, P167.

Robinson E.A. and Treitel S. 1980. *Geophysical Signal Analysis*. Prentice-Hall, Inc.

Scales J.A. and Gersztenkorn A. 1987. Robust methods in inverse theory. In:, *Geophysical Imaging, Symposium of the Geophysical Society of Tulsa* (ed. J.A. Scales), pp. 25–50. Society of Exploration Geophysicists.

Scales J.A. and Gersztenkorn A. 1988. Robust methods in inverse theory. *Inverse Problems* **4,** 1071–1091.

Scales J.A., Gersztenkorn A., Treitel S. and Lines L.R. 1988. Robust optimization methods in geophysical inverse theory. 58th SEG meeting, Anaheim, USA, Expanded Abstracts, 827–830.

Spitz S. 1999. Pattern recognition, spatial predictability, and subtraction of multiple events. *The Leading Edge* **18,** 55–58.

Tarantola A. 1987. *Inverse Problem Theory*. Elsevier Science Publishing Co.

Verschuur D.J. and Berkhout A.J. 1997. Estimation of multiple scattering by iterative inversion, part II: Practical aspects and examples. *Geophysics* **62,** 1596–1611.

Verschuur D.J., Berkhout A.J. and Wapenaar C.P.A. 1992. Adaptive surface-related multiple elimination. *Geophysics* **57,** 1166–1177.

Zhang Z., Chunduru R.K. and Jervis M.A. 2000. Determining bed boundaries from inversion of EM logging data using general measures of model structure and data misfit. *Geophysics* **65,** 76–82.