*Stanford Exploration Project, Report 103, April 27, 2000, pages 123–195*

# Prestack multiple attenuation using the hyperbolic Radon transform

## A comparison of inversion schemes

*Antoine Guitton*[1]

**ABSTRACT**

I apply the iterative hyperbolic Radon transform to CMP gathers to create a velocity panel where multiples and primaries are separable. The velocity panel is created using three different inversion schemes: (1) $l^2$ norm inversion, (2) $l^1$ norm inversion and (3) $l^1$ norm with $l^1$ regularization inversion. The third technique is particularly efficient at separating primaries and multiples in the prestack domain. A comparison of the three techniques shows that some noticeable differences appear in the prestack domain after multiple attenuation and that no discrepancies emerge on the stacked sections. These conclusions are linked to convergence properties of each method, and also linked to the "quality" of the data.

## INTRODUCTION

The last decade has seen an exponential growth in the use of 3-D seismic imaging. Contemporaneous with this development, imaging techniques have become more complex in the effort to account for multi-pathing in complex media and to produce "true amplitude" migrated pictures of the subsurface. Since multiples are not accounted for in the physical model that leads to these migration methods, they can severely affect the final migration result producing erroneous interfaces or amplitude artifacts; consequently, the multiples have to be removed from the data. As pointed out by Weglein (1999), the multiple attenuation techniques may be divided into two families: (1) filtering methods which exploit the periodicity and the separability (move-out discrepancies) of the multiples and (2) the wavefield prediction/subtraction methods, where the multiples are first predicted (Verschhur et al., 1992; Weglein et al., 1997) and then subtracted (Spitz, 1999; Doicin and Spitz, 1991; Dragoset and MacKay, 1993; Clapp and Brown, 1999; Brown et al., 1999).

As oil companies lead exploration towards more complex geological structures (e.g., salt plays) and use 3-D surveys intensively, the attenuation of the multiples becomes more challenging. Spitz (2000, Personal communication) recently asserted that multiples are the number one problem in seismic processing. Traditionally, filtering techniques are the method of choice for multiple processing because of their robustness and cost. However, they have some

[1]**email:** antoine@sep.stanford.edu

limitations when tackling multiples in complex media (predictive deconvolution) and in the preservation of primaries'amplitude ( $f$-$k$ filters). Wavefield methods overcome these limitations, therefore they are becoming more popular in the seismic industry. Nonetheless, they are often arduous to tune, generally slow, and very difficult to extend in 3-D for coverage reasons.

### My approach

This paper will describe my production of a velocity domain where multiples and primaries are easily identifiable and separable by the use of the Hyperbolic Radon Transform (HRT) and inverse theory. The **Huber function** (Huber, 1973), or Huber norm, allows us to solve hybrid $l^1$-$l^2$ inverse problems in an efficient fashion. I compare three different methods to obtain the velocity panel: (1) least-squares inversion, (2) $l^1$ inversion, and (3) $l^1$ inversion with $l^1$ regularization. These velocity panels are then used to perform the multiple suppression (Lumley et al., 1995).

In this paper I will first review the theory of the velocity transform operator. Next, I introduce the Huber norm and the inverse problem I intend to solve to produce the velocity field. Finally, I apply a multiple attenuation technique for different inverse problems to a complete 2-D data set (Mobil AVO data). I will show that the multiple reflections are favorably attenuated with no noticeable differences between the different inverse problems.

## THEORY

### Definitions of operators

The HRT maps the data ($t$, $x$) into a velocity space ($\tau$,$v$) that clearly exhibits the moveout inherent in the data and, therefore, forms a convenient basis for velocity analysis. Thorson and Claerbout (1985) were the first to define the forward and adjoint operators of the HRT, formulating it as an inverse problem, where the velocity domain is the unknown space. In their approach, the forward operator **H** stretches the model space (velocity domain) into the data space (CMP gathers) using a hyperbola superposition principle, whereas the adjoint operator $\mathbf{H}^\dagger$, the HRT, squeezes the data summing over hyperbolas (related to the velocity stack as defined by Taner and Koehler (1969)). The forward operation is

$$d(t,x) = \sum_{s=s_{min}}^{s_{max}} w_o m(\tau = \sqrt{t^2 - s^2x^2}, s), \tag{1}$$

and the adjoint transformation becomes

$$m(\tau,s) = \sum_{x=x_{min}}^{x_{max}} w_o d(t = \sqrt{\tau^2 + s^2x^2}, x), \tag{2}$$

where $x$ is the offset, $s$ the slowness, $\tau$ the two-times zero offset travel time, and $w_o$ a weighting function that compensates to some extent for geometrical spreading and other effects (Claerbout and Black, 1997).

**The inverse problem**

Having defined the forward operator $\mathbf{H}$ and its adjoint $\mathbf{H}^\dagger$, we can now pose the inverse problem. Inverse theory helps us to find a velocity panel which synthesizes a given CMP gather *via* the operator $\mathbf{H}$. In equations, given data $\mathbf{d}$ (CMP gather), we want to solve for the model $\mathbf{m}$ (velocity panel)

$$\mathbf{Hm} = \mathbf{d},$$

which is equivalent to the linear system

$$\mathbf{H}^\dagger\mathbf{Hm} = \mathbf{H}^\dagger\mathbf{d}.$$

This system is easy to solve if $\mathbf{H}^\dagger\mathbf{H} = \mathbf{I}$, i.e., if $\mathbf{H}$ is unitary. Unfortunately, $\mathbf{H}$ is far from an unitary operator. Sacchi and Ulrych (1995) give a couple of reasons for this behavior (see also Kabir and Marfurt (1999) for a more graphical interpretation of the artifacts):

1. The velocity (slowness) range is not wide enough.

2. The sampling in the velocity domain is too coarse.

Inverse problems are often used to handle the non-unitarity of operators. A prior advisable step in the design of an inverse problem is to attribute some properties to the model in terms of moments of corresponding distributions. A reasonable property of the model space would be **sparseness**, meaning that we want to cluster the components of the solution into a few large peaks (Thorson and Claerbout, 1985). The sparseness would help to distinguish primaries and multiples in the velocity space. Finally, we would like to design a solver that bears robustness to bad (inconsistent) data points. These bad data points leave large values in the residual and attract most of the solver's efforts (Fomel and Claerbout, 1995). Unfortunately, seismic data are generally very noisy, and the need for robust estimators is very pressing.

**How to obtain a sparse model?**

Some authors have proposed different solutions to address the sparseness of the model space. Thorson and Claerbout (1985) developed a stochastic inversion scheme that converges to a solution with minimum entropy. Sacchi and Ulrych (1995) apply a very similar method with more degrees of freedom to the choice of parameters. Nichols (1994) uses a regularization term with the $l^1$ norm. All these methods assign long-tailed density functions to the model parameters. Figure 1 shows an exponential (related to the $l^1$ norm) and a Gaussian distribution (related to the $l^2$ norm). The Gaussian distribution will tend to smooth the model space, spreading the energy, whereas the exponential distribution will tend to focus the energy on a few peaks, neglecting average values, and thus leading to a sparse model.

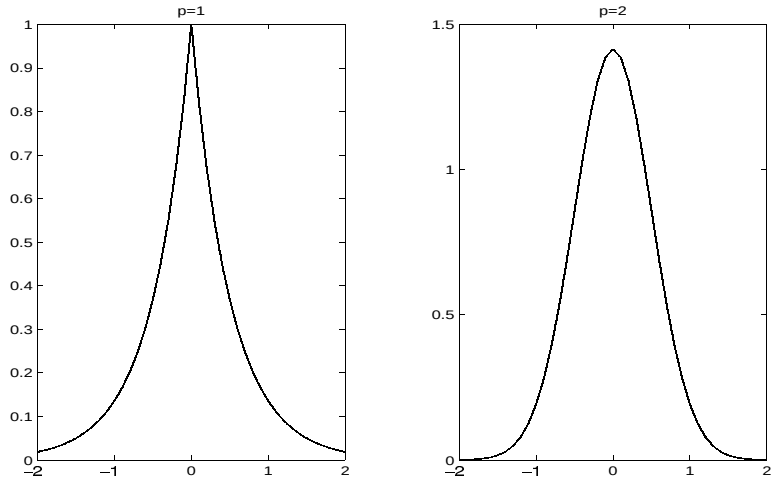


Figure 1: Exponential (left) and Gaussian (right) distribution with zero mean. The exponential distribution has the longer tail. antoine1-distri [CR]

**How to design a robust solver**

The need for a robust solver may be addressed using the $l^1$ norm for the data residual (Claerbout and Muir, 1973). Again, robust measures are related to the long-tailed density function in the same way that the mean square is related to the (short-tailed) Gaussian (Tarantola, 1987). The $l^1$ norm is then less sensitive to outliers and will give a more probable fitting of the data.

The requirements in the design of a robust inverse method that gives a sparse model for the velocity estimation problem leads to the minimization of the objective function

$$f(\mathbf{m}) = |\mathbf{Hm} - \mathbf{d}|_1 + \sigma |\mathbf{m}|_1, \tag{3}$$

where $||_1$ is the $l^1$ norm. Since we wish to utilize the $l^1$ norm, the minimization of $f$ is a cumbersome problem. The $l^1$ norm is not differentiable everywhere, which makes its use rather difficult. The next section presents some alternatives to the $l^1$ norm using hybrid $l^1$-$l^2$ objective functions. These functions are differentiable and allow the use of iterative methods.

**Hybrid $l^1$-$l^2$ function**

The hybrid $l^1$-$l^2$ norm has been widely used in geophysics (Bube and Langan, 1997; Nichols, 1994; Fomel and Claerbout, 1995). It is generally solved using iteratively reweighted least-squares (IRLS) algorithms with an appropriate weighting matrix. These algorithms have proved efficient but are also acknowledged to be difficult to tune. As an alternative to using IRLS algorithms to compute the hybrid $l^1$-$l^2$ norm, Claerbout (1996) introduced the **Huber norm** (Huber, 1973). This norm is a patching of the $l^1$ norm for high residuals and of the $l^2$

norm for small residuals:

$$M_\epsilon(r) = \begin{cases} \frac{r^2}{2\epsilon}, & 0 \le |r| \le \epsilon \\ |r| - \frac{\epsilon}{2}, & \epsilon < |r| \end{cases} \tag{4}$$

where $r$ is the residual. We call $\sum_{i=1}^{N} M_\epsilon(r_i)$ the *Huber misfit function*, or the Huber function, for short (Figure 2). Note that the Huber function is smooth near zero residual and weights small residuals by the mean square. It is reasonable to suppose that the Huber function, while maintaining robustness against large residuals, is easier to minimize than $l^1$. The parameter $\epsilon$, which controls the limit between $l^1$ and $l^2$, is called the Huber threshold.

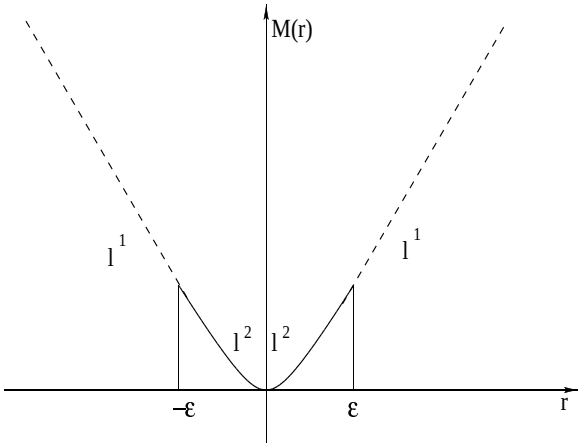


Figure 2: Error measure proposed by Huber (1973). The upper part above $\epsilon$ is the $l^1$ norm, while the lower part is the $l^2$ norm. antoine1-huber [NR]

The implementation of an inverse solver to minimize the Huber function is quite challenging and leads to innovative non-linear algorithms in geophysics (Guitton, 2000b). To summarize, I developed a quasi-Newton method with a line search. I implemented a More and Thuente Line Search (More and Thuente, 1994) algorithm, which ensures a sufficient decrease in the objective function $f$ (equation 3) and obeys curvature conditions (the so-called *Wolfe conditions*, Kelley (1999)). The update of the Hessian is made using a Limited Memory BFGS method as proposed by Nocedal (1980) and Liu and Nocedal (1989). This method is guaranteed to converge to a minimum. This strategy has proved efficient in solving the Huber problem correctly (Guitton and Symes, 1999; Guitton, 2000a) and eliminates the restart parameter encountered in IRLS algorithms, which makes this *Huber solver* easier to use.

**What will I do?**

My goal in this paper is to compare three different inversion schemes for the multiples attenuation problem. They all aim to produce a velocity model where primaries are muted out and the predicted multiples are subtracted from the original data. I successively solve

1. $f(\mathbf{m}) = \|\mathbf{Hm} - \mathbf{d}\|_2$,
2. $f(\mathbf{m}) = |\mathbf{Hm} - \mathbf{d}|_1$,

3. $f(\mathbf{m}) = |\mathbf{Hm} - \mathbf{d}|_1 + \sigma |\mathbf{m}|_1$,

and compare the results. I call arbitrarily "$l^1$ norm" any Huber function with a small threshold. Let us assume now that to the $l^1$ norm, for the data residual, corresponds a threshold

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

In addition, for the regularization term, let us say that to the $l^1$ norm corresponds a threshold

$$\epsilon = \frac{max|\mathbf{d}|}{10000}.$$

$\epsilon$ is chosen smaller than before leading to a larger $l^1$ treatment of the model. I show later on that the convergence is greatly reduced by the addition of this regularization term.

## MARINE DATA RESULTS

This section presents results obtained using the HRT on a the Mobil AVO data set (Lumley et al., 1995). These prestack data are heavily contaminated by free-surface and water-bottom multiples and thus constitute a challenging test-bed for true amplitude multiple attenuation techniques. The strategy for the multiple elimination is as follows:

1. Create a velocity panel with the iterative Hyperbolic Radon Transform using the Huber solver.
2. Define a corridor between multiples and primaries in the velocity space.
3. Mute out the primaries in the velocity space.
4. Model back the multiples in the data space.
5. Subtract the multiples field from the input data.
6. NMO and Stack.

This method is most suitable for preserving primaries energy. I compare $l^2$, $l^1$ and $l^1$ with $l^1$ regularization for the inversion in step 1. I decide to parameterize the problem using velocities and not slownesses.

### Computing aspects

The velocity-stack inversion has been fully automated, implying that a threshold is adaptively computed for each CMP gather independently. The input data are composed of 801 gathers

with 60 traces each. In my implementation, I parallelized the computing using OpenMP [2](this concerns the operators only). On 10 processors of our SGI Power Challenge machine (notoriously slow), for 30 iterations, it took 30 hours to compute the 801 velocity panels. This process may be accelerated by faster machines.

**Inversion results**

Figure 3 shows the result of the inversion for one CMP. Since the Mobil AVO dataset does not include very complex structures with strong velocity contrasts, this panel illustrates what happens for all the gathers. The left panel shows the input data. The other panels display the reconstructed data using the different schemes. Note that the $l^2$ and $l^1$ inversions give similar results and that the $l^1$ regularization doesn't converge as well. Figure 4 highlights this difference between the different problems. The best convergence is achieved using least-squares and the worst is achieved with the $l^1$ regularization. In my implementation, however, the $l^1$ problem with or without regularization was solved using twice as many iterations as with $l^2$. Figure 5 shows the differences between the input data and the remodeled data. It appears that the $l^1$ norm with $l^1$ regularization encounters some difficulties in fitting the far offset data. Note that the $l^1$ norm and the $l^2$ norm are both comparable. This is expected since the data are not strongly noisy.

Differences arise in favor of the $l^1$ regularization when we look at the model space (Figure 6), however. The $l^1$ and $l^2$ results are again very similar and the $l^1$ norm with $l^1$ regularization appears spikier. This result is consistent with the theory (see Theory section). The spiky result is then used to define the limit between primaries and multiples (black line in the right panel of Figure 6). A mask is defined accordingly and the primaries are muted out in the model space. The next step consists of remodeling the multiples back in the data space, applying the hyperbola superposition principle (operator **H**). Figure 7 shows the predicted multiples. Note that for the three inversion schemes, some primaries remain. This is particularly annoying to us in our attempt to produce true amplitude multiple-free gathers. Figure 8 displays the result of the multiple attenuation process. The three methods display similar results. Nonetheless, at far offset, the $l^1$ regularization shows more energetic events. This is consistent with Figure 5 where we showed that the $l^1$ regularization was unable to fit this part of the data.

**NMO-Stacking process**

In the last step, the data are NMO corrected and stacked. The stacking velocity comes from Lumley et al. (1995). Figure 9 shows the stack of the input data without multiple attenuation. Figures 10, 12, 14 show the stacks of the multiple-free data with the different methods. Figures 11, 13, 15 show the difference between the stacked section of the input data *with* multiples and the stacked section of the data *without* multiples for the three inversion schemes. We see that the multiple suppression has cleaned up deeper parts of the structure (below 2s.). Furthermore,

[2]OpenMP is a specification for a set of compiler directives, library routines, and environment variables that can be used to specify shared memory parallelism in Fortran and C/C++ programs (www.openmp.org).

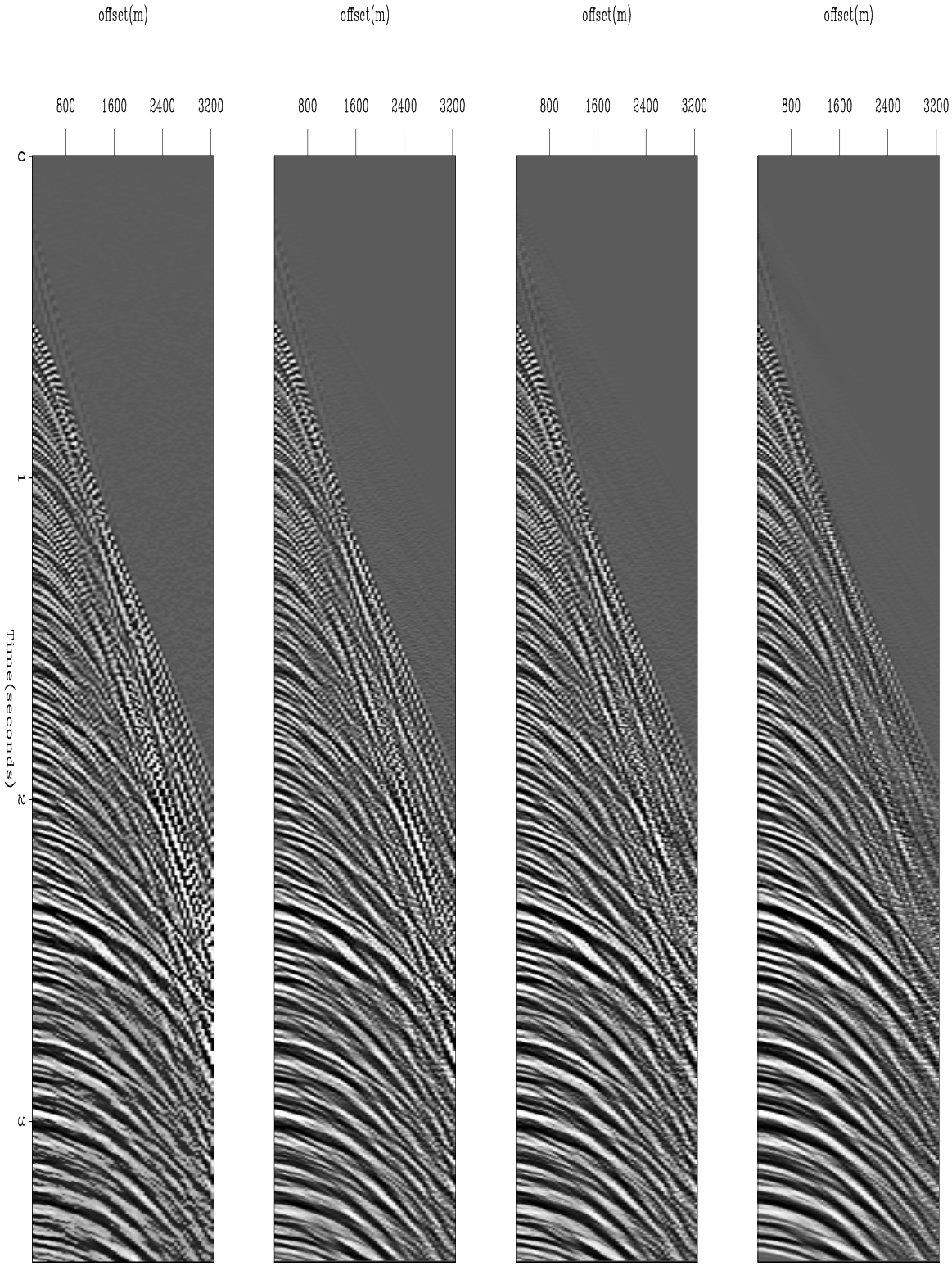


Figure 3: Left: input data. Middle-left: $l^2$ reconstructed data. Middle-right: $l^1$ reconstructed data. Right: $l^1$ with $l^1$ regularization reconstructed data. antoine1-comp_dat [CR]

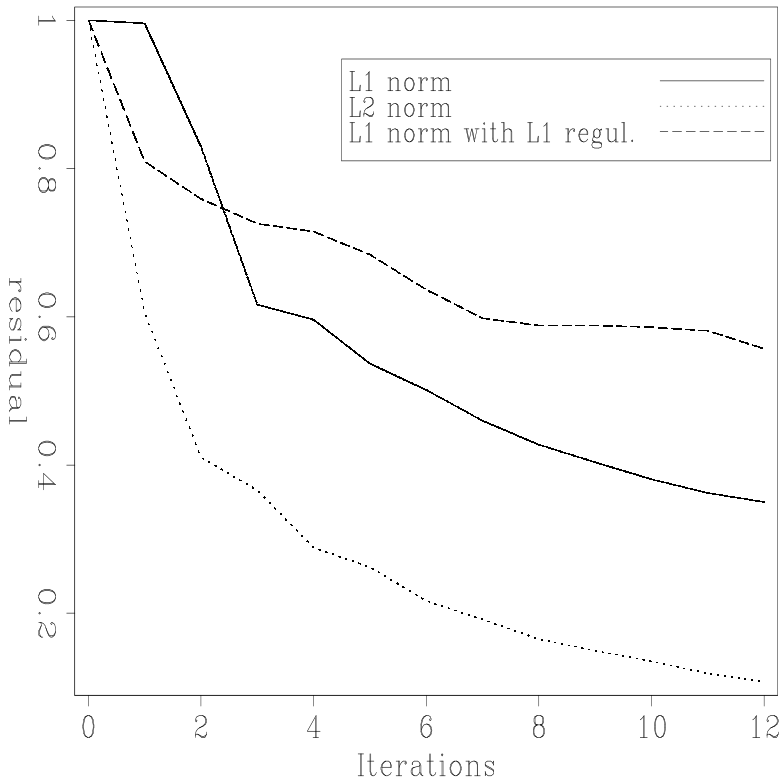


Figure 4: Comparison of the convergence for different inversion schemes for one CMP gather. antoine1-residual [CR]

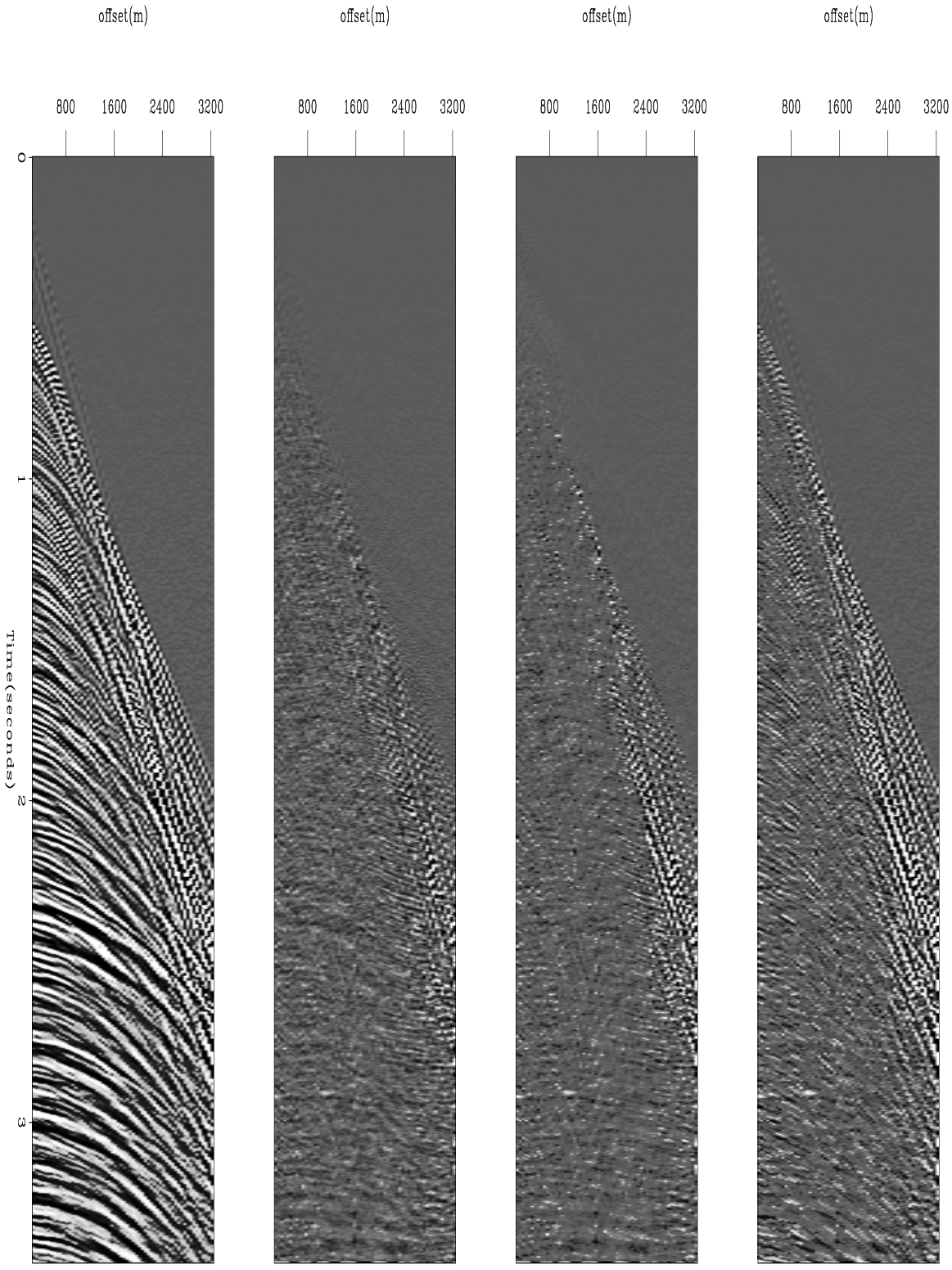


Figure 5: Left: input data. Middle-left: $l^2$ residual. Middle-right: $l^1$ residual. Right: $l^1$ with $l^1$ regularization residual. antoine1-diff [CR]

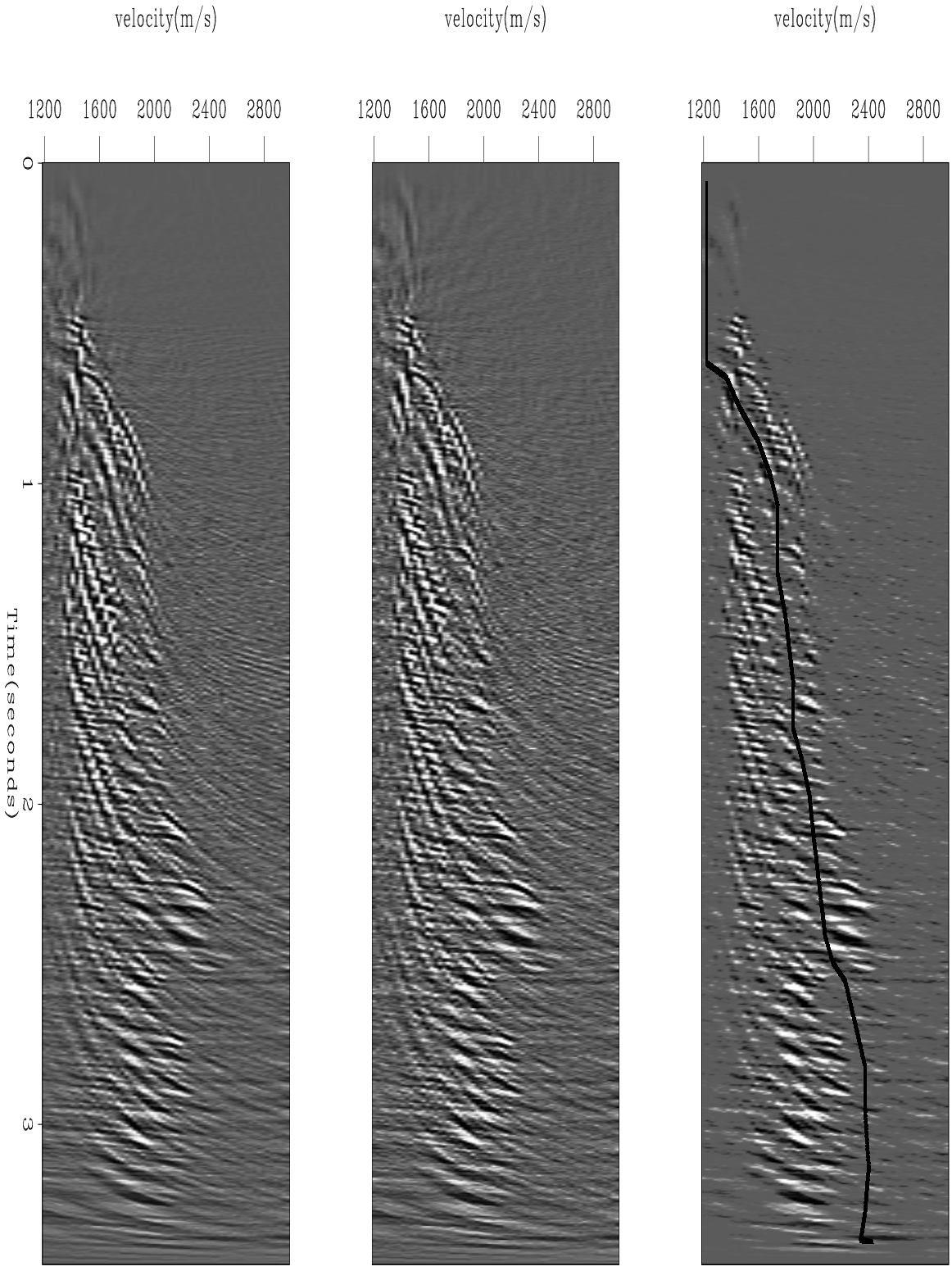


Figure 6: Left:$l^2$ model. Middle: $l^1$ model. Right: $l^1$ with $l^1$ regularization. The line shows the limit of the muting process that separates "guessed" multiples on the left from "guessed" primaries on the right for the spiky model. antoine1-comp_scan [CR]

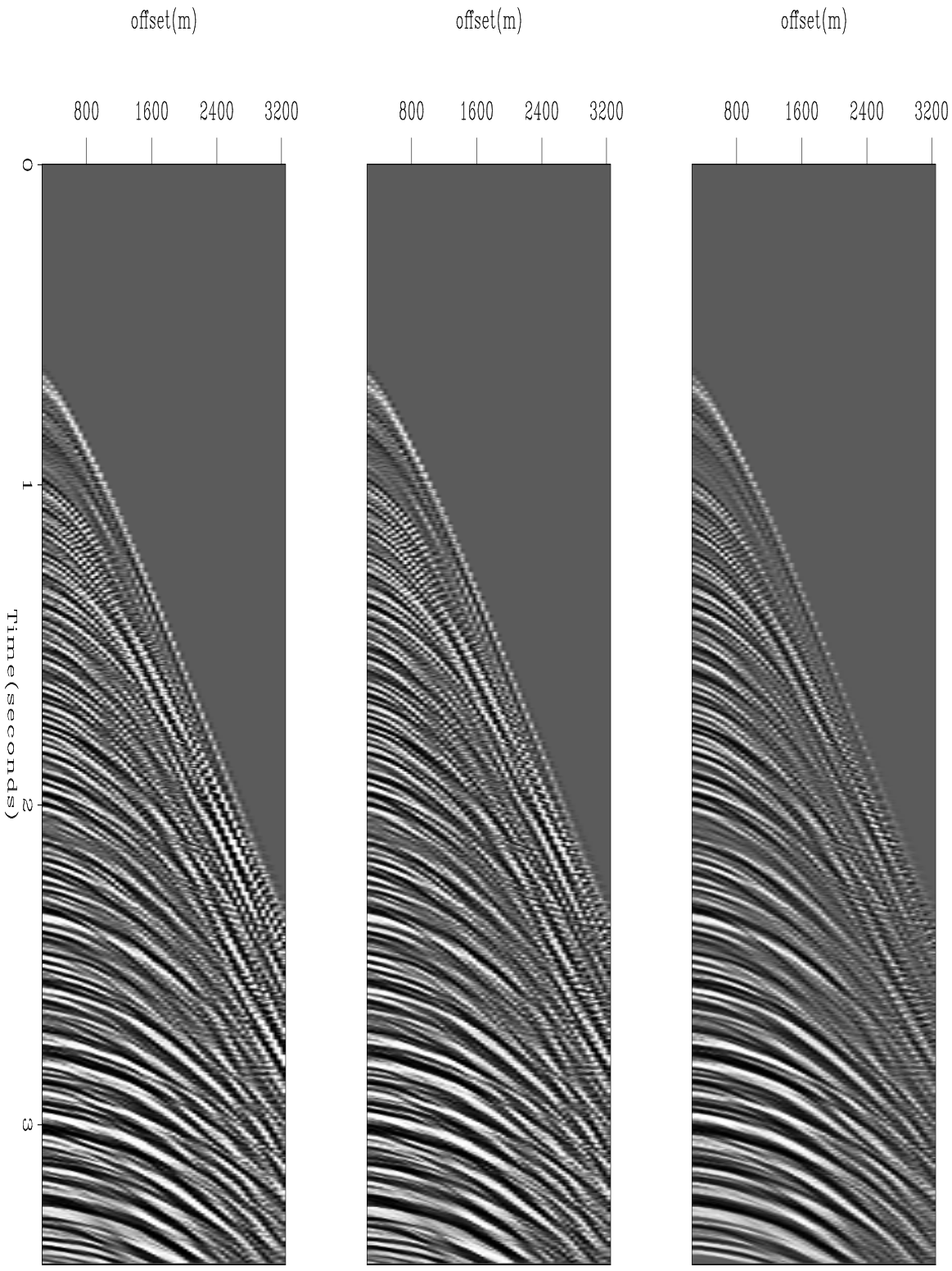


Figure 7: Predicted multiples. Left: $l^2$ multiples. Middle: $l^1$ multiples. Right: $l^1$ with $l^1$ regularization multiples. antoine1-comp_mult [CR]

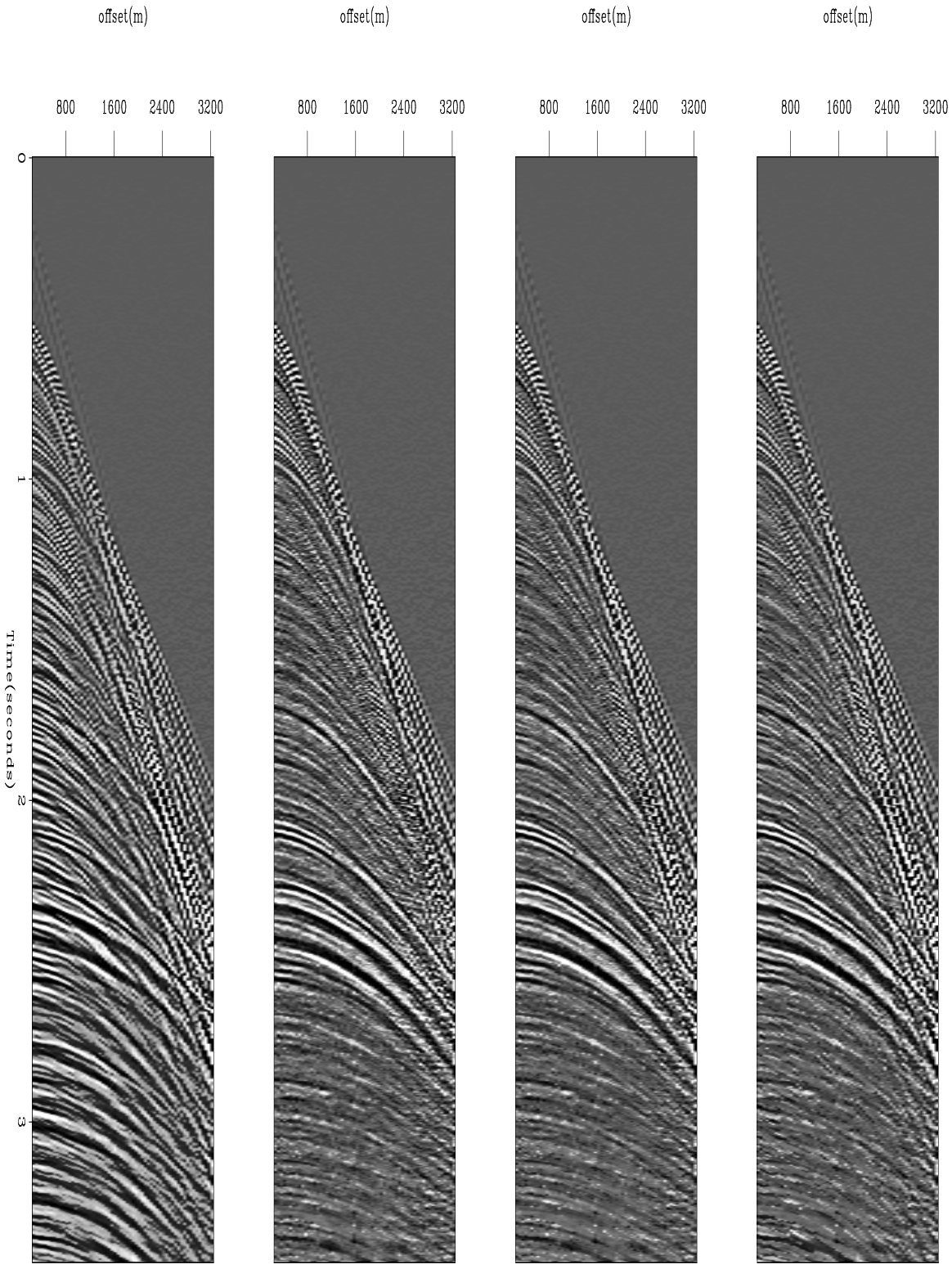


Figure 8: Gathers after multiple attenuation. Left: input data with multiples. Middle-left: $l^2$ multiple attenuation. Middle-right: $l^1$ multiple attenuation. Right: $l^1$ with $l^1$ regularization multiple attenuation. antoine1-comp [CR]

there are no noticeable differences between the three methods in the stacked sections. In addition, as is often the case, the stacking is such a powerful multiple-suppression method that it attenuates multiples sufficiently well: the simple stack of the input data with multiples looks fairly close to the stacked section without multiples (see Figure 9).

### Discussion

The overall results may seem disappointing: the simple stacking of the input data give similar results to the stacking of multiple-free gathers using $l^2$, $l^1$ and $l^1$ with $l^1$ regularization. Nonetheless, data with more complex multiples would certainly lead to different conclusions. In addition, these data demonstrate that when the gathers are not particularly noisy, which is the case here, the $l^1$ norm and the $l^2$ norm behave similarly. The $l^1$ norm with $l^1$ regularization produces expected sparse velocity panels with very bad convergence properties, meaning that we need to think about new strategies to improve it. Claerbout (2000, Personal communication) recently suggested that I minimize

$$f(\mathbf{m}) = \|\mathbf{R}(\mathbf{Hm} - \mathbf{d})\|$$

where **R** is the Prediction Error Filter (PEF) of the residual. The PEF would help to obtain Independent Identically Distributed (IID) variables in the residual. Another idea is to minimize

$$f(\mathbf{m}) = |\mathbf{F}.\mathbf{T}_{\mathbf{2-D}}(\mathbf{Hm} - \mathbf{d})|_{\mathbf{Huber}}$$

where $F.T_{2-D}$ is the 2-D Fourier Transform. The idea behind this last equation is that the far offset data, which the $l^1$ norm with $l^1$ regularization does not fit very well, creates (almost) mono-frequency patterns in the residual that would map in a very localized area of the Fourier space with high amplitudes (two points for a perfectly mono-frequency event with one slope). The Huber norm with an appropriate threshold would treat these focused energies as outliers and get rid of them.

## CONCLUSION

I have shown that the multiple reflections may be attenuated in the prestack domain using the Hyperbolic Radon Transform. Different inverse problems have been solved to obtain the velocity panels showing different properties: (1) the $l^2$ and $l^1$ norm produce comparable results in the velocity space and for the reconstructed data, (2) the $l^1$ norm with $l^1$ regularization shows spikier results in the velocity domain but converge much more slowly than the other methods, and (3) the stacked sections of the multiple-free data are very similar for the different inverse problems. The good quality of the data explains the small discrepancy between the $l^2$ and the $l^1$ norm; more noisy gathers would lead to different conclusions.

## REFERENCES

Brown, M., Clapp, B., and Marfurt, K., 1999, Predictive signal/noise separation of ground-roll contaminated data: SEP–**102**, 111–128.

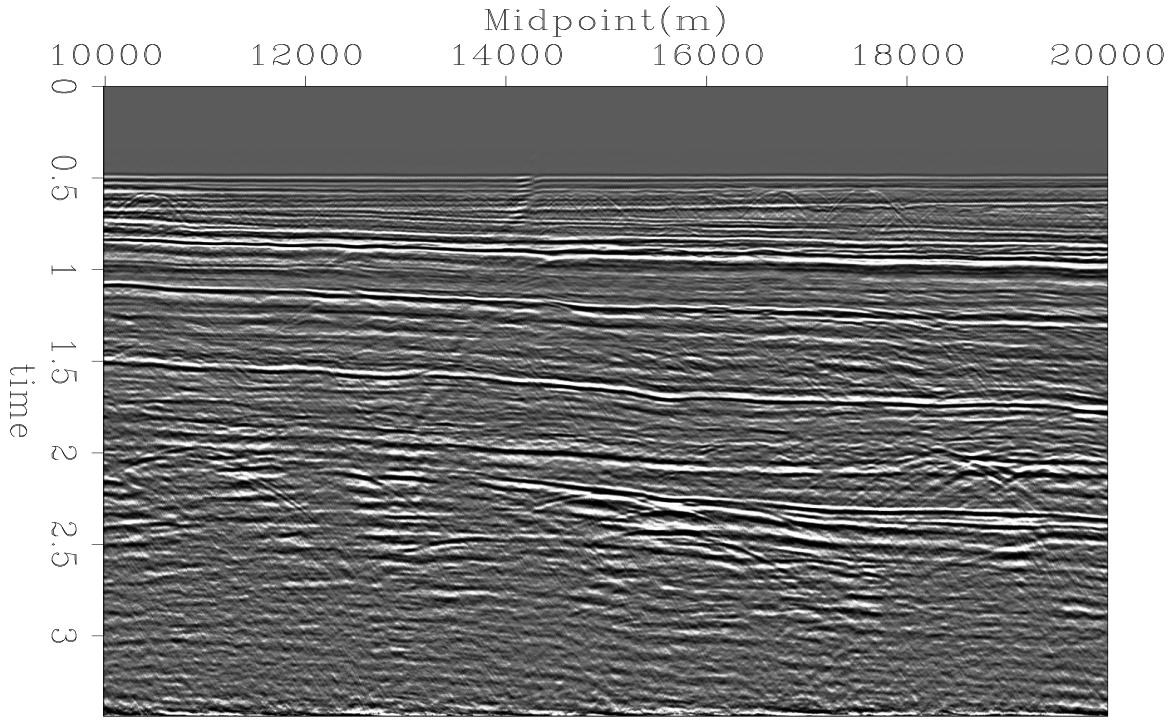


Figure 9: Stacked section of the input data with multiples. antoine1-stack_IN [CR]

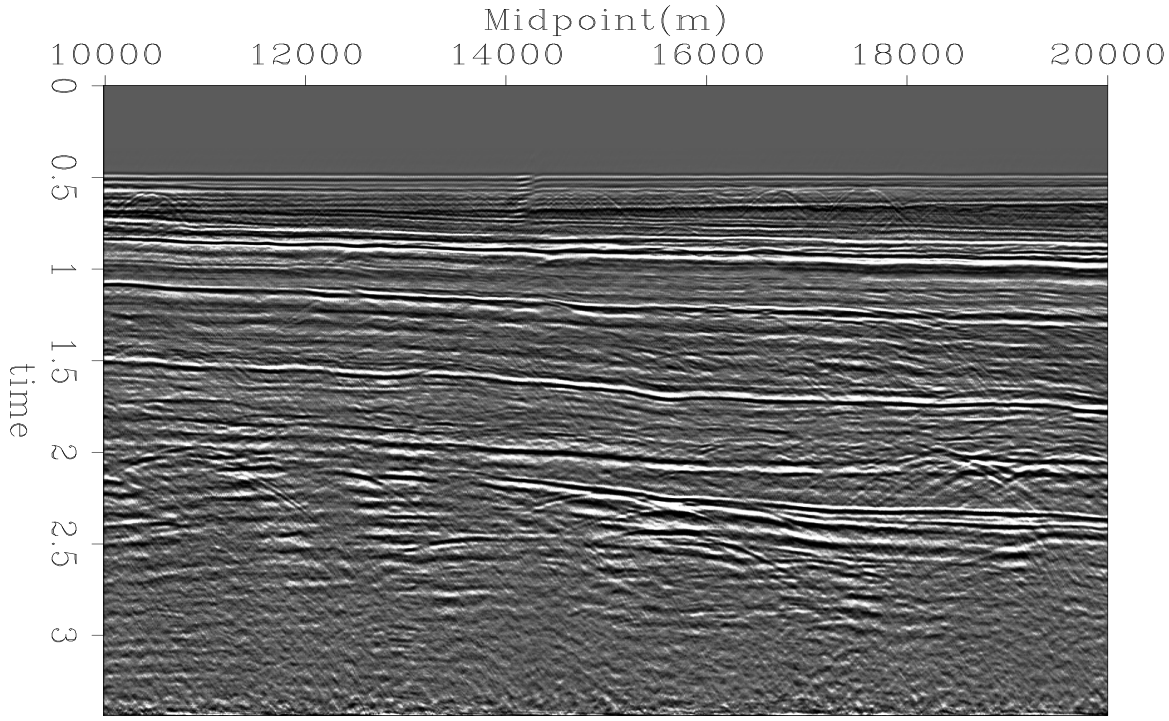


Figure 10: Stacked section after multiple suppression using the $l^2$ norm. antoine1-stack_L2 [CR]

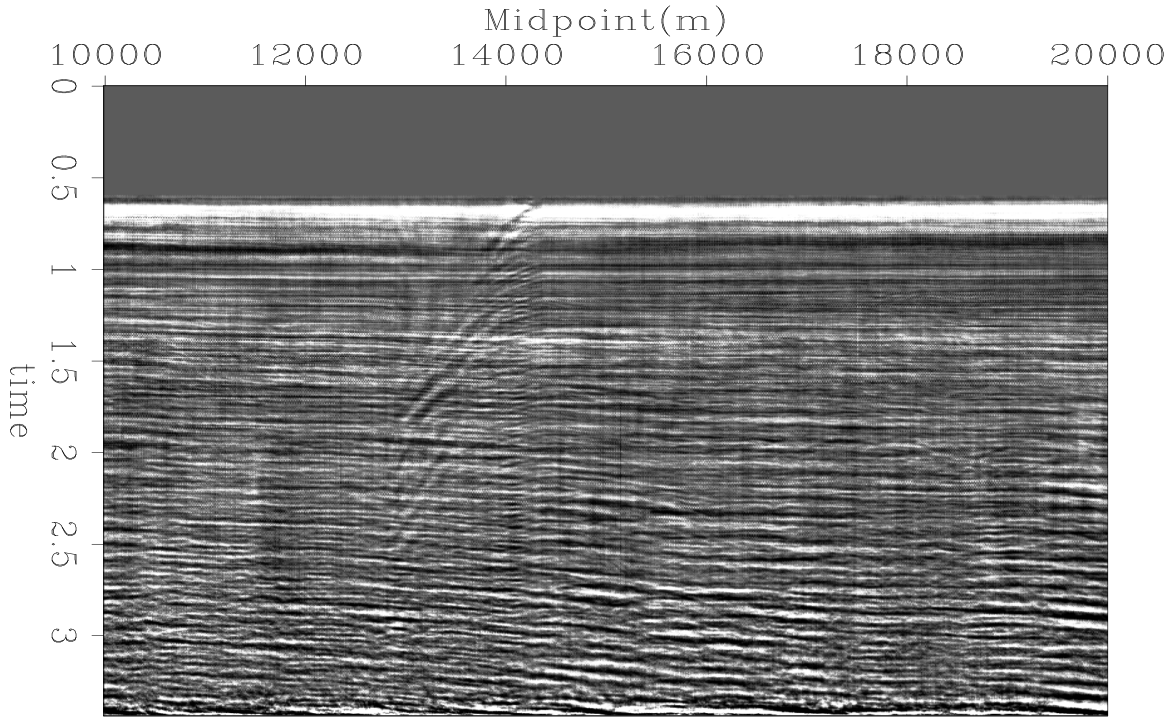


Figure 11: Difference between the stacked section with multiples and the stacked section without multiples using the $l^2$ norm. antoine1-comp_stack_L2 [CR]

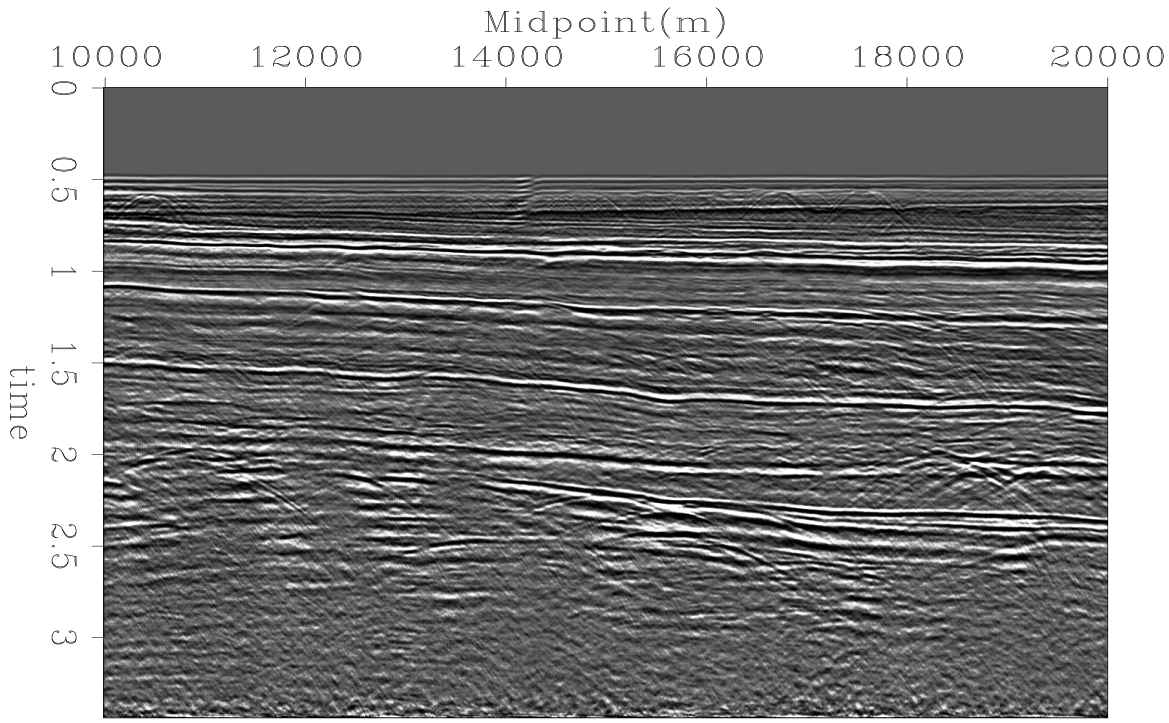


Figure 12: Stacked section after multiple suppression using the $l^1$ norm. antoine1-stack_L1 [CR]

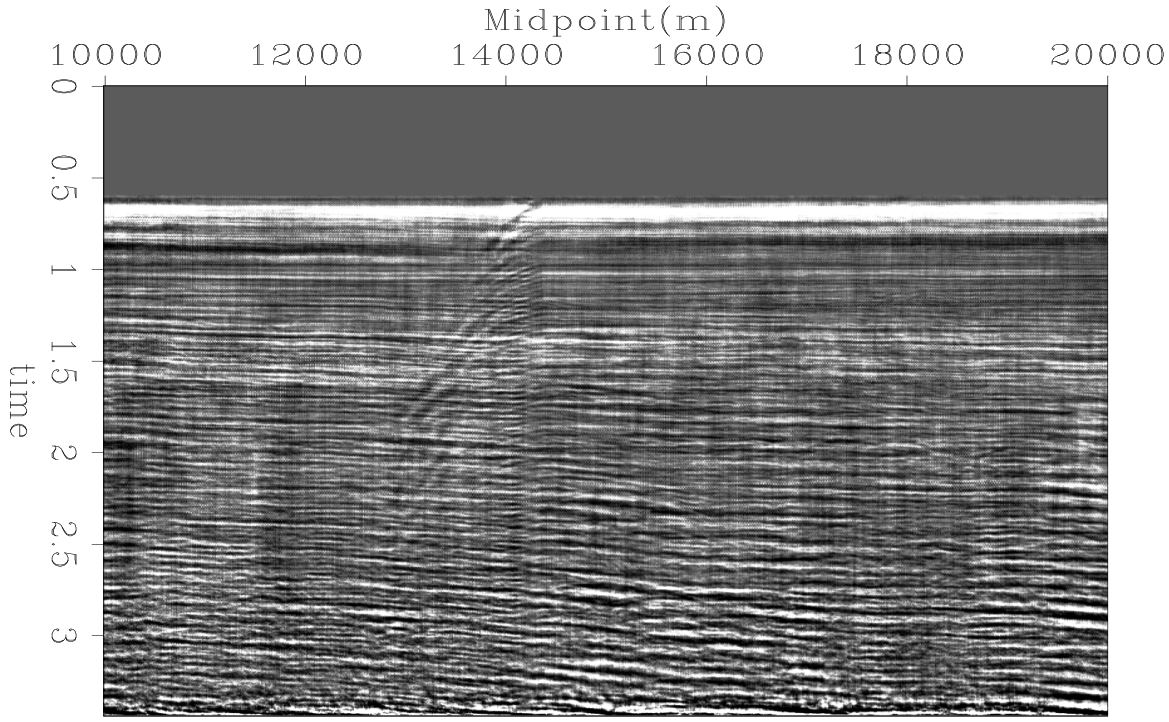


Figure 13: Difference between the stacked section with multiples and the stacked section without multiples using the $l^1$ norm. antoine1-comp_stack_L1 [CR]

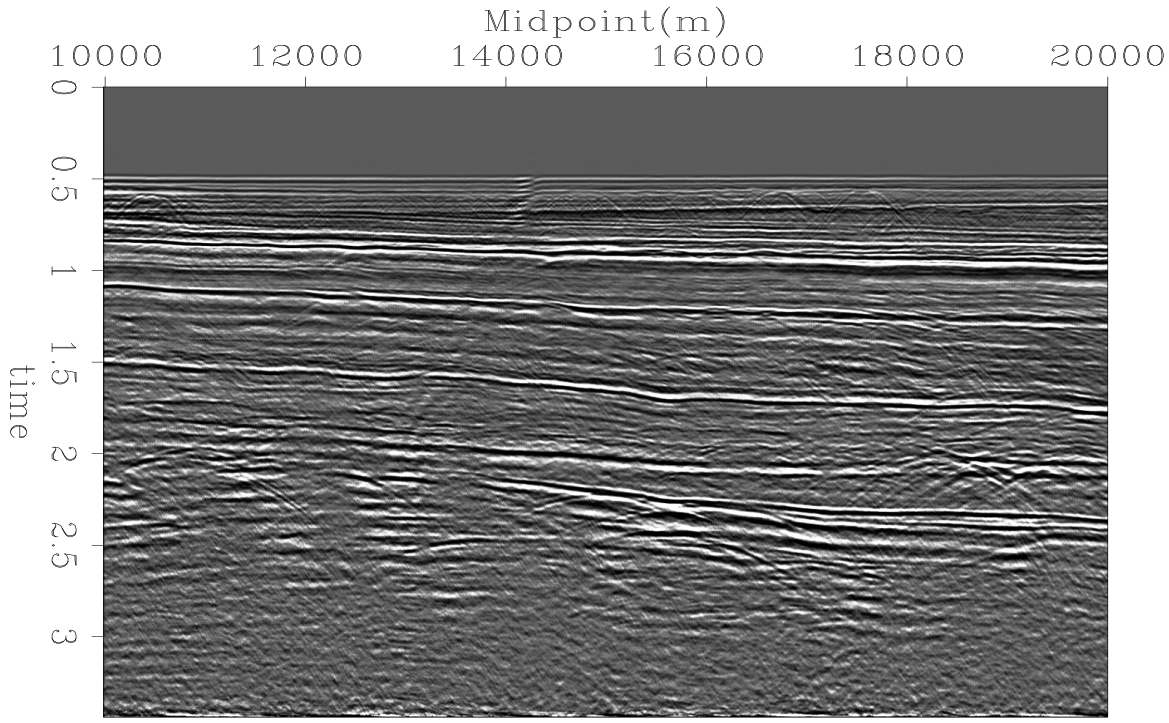


Figure 14: Stacked section after multiple suppression using the $l^1$ norm with $l^1$ regularization. antoine1-stack_L1L1 [CR]

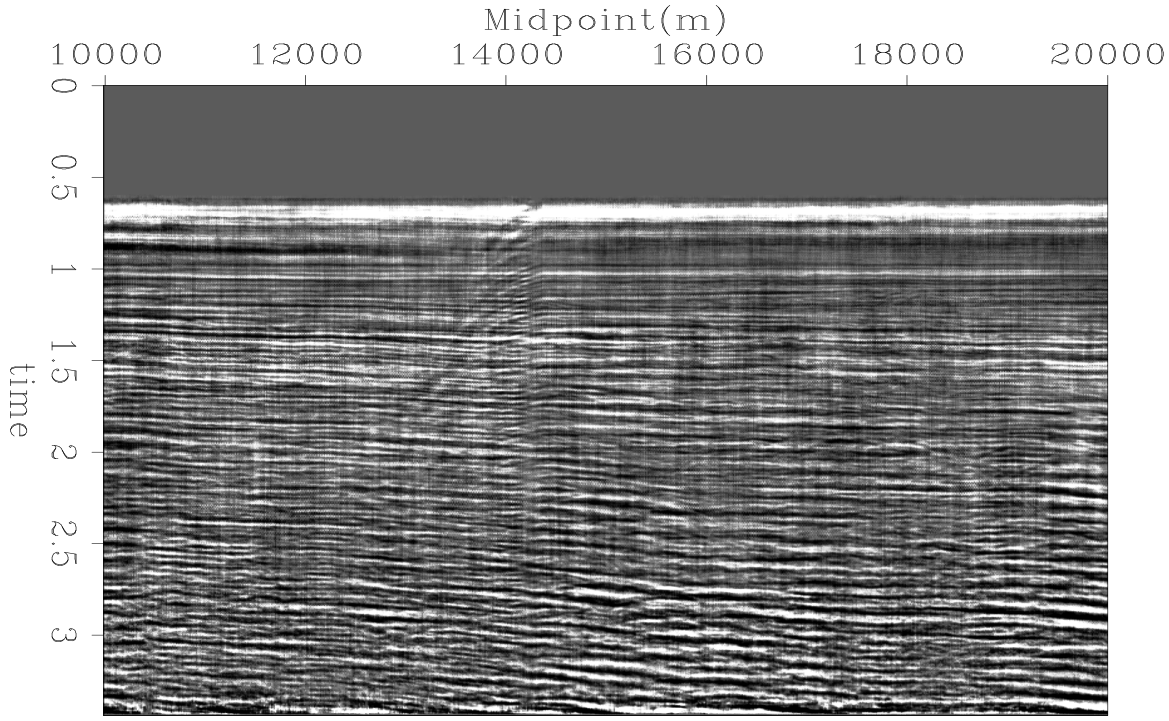


Figure 15: Difference between the stacked section with multiples and the stacked section without multiples using the $l^1$ norm with $l^1$ regularization. antoine1-comp_stack_L1L1 [CR]

Bube, K. P., and Langan, R. T., 1997, Hybrid lambda1/lambda2 minimization with applications to tomography: Geophysics, **62**, no. 04, 1183–1195.

Claerbout, J. F., and Black, J. L., 1997, Basic Earth Imaging: Class notes, http://sepwww.stanford.edu/sep/prof/index.html.

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

Claerbout, J., 1996, Conjugate-direction Huber regression: SEP–**92**, 229–235.

Clapp, B., and Brown, M., 1999, Applying SEP's latest tricks to the multiple suppression problem: SEP–**102**, 91–100.

Doicin, D., and Spitz, S., 1991, Multichannel extraction of water bottom peg legs pertaining to a high amplitude reflection: 61th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1439–1442.

Dragoset, B., and MacKay, S., 1993, Surface multiple attenuation and subsalt imaging: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1099–1102.

Fomel, S., and Claerbout, J. F., 1995, Searching the sea of galilee: SEP–**94**, 259–270.

Guitton, A., and Symes, W. W., 1999, Robust and stable velocity analysis using the Huber function: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1166–1169.

Guitton, A., 2000a, Huber solver versus IRLS algorithm for quasi L1 inversion: SEP–**103**, 255–271.

Guitton, A., 2000b, Implementation of a nonlinear solver for minimizing the Huber norm: SEP–**103**, 281–289.

Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: Ann. Statist., **1**, 799–821.

Kabir, N. M. M., and Marfurt, K. J., 1999, Toward true amplitude multiple removal: The Leading Edge, **18**, 66–73.

Kelley, C. T., 1999, Iterative methods for optimization: SIAM in applied mathematics.

Liu, D. C., and Nocedal, J., 1989, On the limited memory BFGS method for large-scale optimization: Mathematical Programming, **45**, 503–528.

Lumley, D., Nichols, D., Ecker, C., Rekdal, T., and Berlioux, A., 1995, Amplitude-preserved processing and analysis of the mobil avo data set: SEP–**84**, 125–152.

More, J. J., and Thuente, J., 1994, Line search algorithms with guaranteed sufficient decrease: ACM Transactions on Mathematical Software, **20**, 286–307.

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

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **95**, 339–353.

Sacchi, M. D., and Ulrych, T. J., 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, no. 04, 1169–1177.

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

Taner, M. T., and Koehler, F., 1969, Velocity spectra-digital computer derivation and applications of velocity functions: Geophysics, **34**, no. 06, 859–881.

Tarantola, A., 1987, Inverse Problem Theory: Elsevier Science Publisher.

Thorson, J. R., and Claerbout, J. F., 1985, Velocity stack and slant stochastic inversion: Geophysics, **50**, no. 12, 2727–2741.

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

Weglein, A. B., Gasparotto, F. A., Carvalho, P. M., and Stolt, R. H., 1997, An inverse scatering series method for attenuating multiples in seismic reflection data: Geophysics, **62**, 1975–1989.

Weglein, A. B., 1999, Multiple attenuation: an overview of recent advances and the road ahead (1999): The Leading Edge, **18**, 40–44.