

## INVERSION OF PRESTACK SEISMIC DATA USING FISTA

Daniel O. Pérez<sup>a</sup>, Danilo R. Velis<sup>a</sup> and Mauricio D. Sacchi<sup>b</sup>

<sup>a</sup>*Facultad de Ciencias Astronómicas y Geofísicas, UNLP, Paseo del Bosque s/n, 1900 La Plata; and CONICET, Argentina, [dperez@fcaglp.unlp.edu.ar](mailto:dperez@fcaglp.unlp.edu.ar), [velis@fcaglp.unlp.edu.ar](mailto:velis@fcaglp.unlp.edu.ar)*

<sup>b</sup>*Department of Physics, University of Alberta, Edmonton, Canada, [msacchi@ualberta.ca](mailto:msacchi@ualberta.ca)*

**Keywords:** FISTA, inversion,  $l_1$ -norm, prestack data

**Abstract.** In this work we present a new inversion method to obtain AVA high-resolution attributes from prestack seismic data. The method aims to find a series of sparse reflectors that, when convolved with the source wavelet, fit the observed data. To perform the inversion, we propose the use of the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA). FISTA, which can be viewed as an extension of the classical gradient algorithm, provides sparse solutions minimizing both the misfit between the modeled and the observed data, and the  $l_1$ -norm of the solution. The advantage of FISTA over other methods is that no inversion over any matrix is needed, making it numerically stable, easy to apply, economic in computational terms, and adequate for solving large-scale problems even with dense matrix data. Results on synthetic and field data show that the proposed method is capable to provide high-resolution AVA attributes that honor the observed data under noisy conditions, making it an interesting alternative to other known methods.

## 1 INTRODUCTION

One of the objectives of the inversion of prestack seismic data is the determination of contrasts between rock properties such as compressional-wave velocities, shear-wave velocities and densities. These contrasts can be estimated through the analysis of the observed variation of the amplitudes of the reflected waves with the angle of incidence. The amplitude-versus-angle (AVA) variation can be described by the Zoeppritz equations (Zoeppritz, 1919; Yilmaz, 2001), which provide the relationship between the amplitude of the reflected compressional-waves and the angle of incidence for a given plane wave that arrives at an interface that separates two different media. Due to their high non-linearity, Zoeppritz equations are impractical for applications such as data interpretation and inversion. Hence, over the last two decades, various authors have developed several linear approximations (Aki and Richards, 1980; Shuey, 1985; Fatti et al., 1994). The coefficients of such approximations, which are the objective of the inversion, constitute AVA attributes that may provide important information about fluid content, a key issue for the characterization of hydrocarbon reservoirs (Castagna et al., 1998; Smith and Gidlow, 2000).

Due to the non-uniqueness inherent to this type of inverse problem, there might exist several sets of coefficients that honor the data equally well, and because some of these might exhibit a huge  $l_2$ -norm they are meaningless. More useful solutions can be obtained by the use of some kind of regularization to promote a certain type of solution through the minimization of a suitable norm. Sparseness is a property that can be incorporated as *a priori* information through  $l_1$ -norm regularization. Sparse solutions are desirable because they can be used to characterize significant and close reflectors more accurately than using traditional quadratic  $l_2$ -norm regularization. In this sense,  $l_1$ -norm regularization favors sparse-spike solutions that lead to high-resolution images, though at the expense of some mathematical burden and increased computational cost.

Sparse-spike AVA inversion has been studied with very interesting results by several authors, either based on a Bayesian inversion where sparseness is obtained through the use of appropriate long-tailed *a priori* probability distributions (Downton and Lines, 2003; Misra and Sacchi, 2008; Alemie and Sacchi, 2011), or based on global optimization algorithms where sparseness is incorporated as an *a priori* condition (Pérez and Velis, 2011). The method proposed in this paper shares the same objectives as those works, but introduces a simple and cost-effective new procedure to solve this kind of inverse problems: the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA). FISTA is a powerful algorithm presented very recently (Beck and Teboulle, 2009) that can be used to minimize, together with the misfit term, the  $l_1$ -norm; but as far as we know, has never been applied for seismic data inversion. FISTA is based on the Iterative Shrinkage Thresholding Algorithm (ISTA) (Daubechies et al., 2004), an extension of the classical gradient algorithm to solve large-scale linear inverse problems in a simple way, where at each iteration only matrix-vector multiplications, and no matrix inversions, are involved. ISTA is known to have slow convergence, but FISTA is shown to be faster by several orders of magnitude. In practice, both the misfit and the  $l_1$ -norm are combined into a cost function by means of a trade-off parameter  $\lambda$  that balances their overall impact.

We tested the method on synthetic normal-move-out (NMO) corrected prestack data using the classical two-term Shuey's approximation to the Zoeppritz equations (Shuey, 1985), showing that high-resolution AVA attributes can be derived from noisy data very accurately. Tests on field data allowed us to obtain high-resolution AVA attribute images such as *Intercept* and *Gradient* images that honor the observed data and show a good lateral continuity.

## 2 THEORY

For the inversion we rely on the convolutional model. This model assumes that the medium is composed of a series of flat, parallel, homogeneous and isotropic layers. Given an angle-gather, the trace corresponding to the  $i$ -th angle of incidence  $\theta_i$  can be expressed as

$$\mathbf{s}(\theta_i) = \mathbf{w} * \mathbf{r}(\theta_i) + \mathbf{n}(\theta_i), \quad i = 1, \dots, N \quad (1)$$

where  $\mathbf{w}$  is the source wavelet of dimension  $L_w$ ,  $\mathbf{r}(\theta_i)$  is the reflectivity of dimension  $L_r$ , and  $\mathbf{n}(\theta_i)$  is the random noise. Both  $\mathbf{n}(\theta_i)$  and  $\mathbf{s}(\theta_i)$  have dimension  $L_s = L_w + L_r - 1$ .

The reflection coefficient for a reflector at time  $t$  can be approximated, in a general form, as

$$r_t(\theta) = \sum_{k=1}^n x_{tk} g_k(\theta), \quad (2)$$

where  $x_{tk}$  are coefficients that depend on the physical properties of the rocks on each side of the interface (velocities and densities),  $n$  is the order of the chosen approximation (usually  $n = 2$  or  $3$ ), and  $g_k(\theta)$  are functions that depend on the angle of incidence, which must be less than the critical angle (Ikelle and Amundsen, 2005).

Combining equations (1) and (2), and omitting the noise term for simplicity, it yields

$$\mathbf{s}(\theta_i) = \mathbf{A}(\theta_i)\mathbf{x}, \quad i = 1, \dots, N \quad (3)$$

where  $\mathbf{x} = (x_{11}, \dots, x_{L_r,1}, \dots, x_{1n}, \dots, x_{L_r,n})^T$  and  $\mathbf{A}(\theta_i)$  is an augmented matrix of dimension  $L_s \times nL_r$  that can be expressed as

$$\mathbf{A}(\theta_i) = (\mathbf{A}_1(\theta_i) | \dots | \mathbf{A}_n(\theta_i)). \quad (4)$$

Here,  $\mathbf{A}_k(\theta_i)$  with  $k = 1, \dots, n$  are sub-matrices of dimension  $L_s \times L_r$ , whose elements are given by

$$[\mathbf{A}_k(\theta_i)]_{h,j} = g_k(\theta_i) w_{h-j+1}, \quad (5)$$

for  $h = 1, \dots, L_s$  and  $j = 1, \dots, L_r$ . Then, the  $N$  systems of equations given by equation (3) can be arranged in a unique system in the form

$$\mathbf{A}\mathbf{x} = \mathbf{s}, \quad (6)$$

where  $\mathbf{A}$  is a column block matrix with blocks given by  $\mathbf{A}(\theta_i)$ , and  $\mathbf{s}$  is a one column block vector with blocks given by  $\mathbf{s}(\theta_i)$ .

A sparse solution of the system given by equation (6) can be estimated using least-squares with  $l_1$ -norm regularization, which implies to find the  $\mathbf{x}$  that minimizes the cost function

$$J = \|\mathbf{A}\mathbf{x} - \mathbf{s}\|^2 + \lambda \|\mathbf{x}\|_1. \quad (7)$$

In this equation, the first term, which represents the error or misfit, is used to measure the differences between the observed and the modeled data. On the other hand, the second term is used to penalize non-sparse solutions. The trade-off parameter  $\lambda$  is used to balance the weight of the two terms. To find the minimum of  $J$  we propose the use of FISTA (Beck and Teboulle, 2009). Step-by-step, FISTA is as follow:

1. Set the constant  $\alpha$  greater than or equal to the maximum eigenvalue of  $\mathbf{A}^T \mathbf{A}$ .
2. Set  $\mathbf{z}_1 = \mathbf{x}_0$  and  $t_1 = 1$ , where  $\mathbf{x}_0$  is an initial solution.
3. For each FISTA iteration  $k = 1, 2, 3, \dots$ :

(a)

$$\mathbf{x}_k = T_{\lambda/2\alpha} \left\{ \mathbf{z}_k - \frac{1}{\alpha} \mathbf{A}^T (\mathbf{A} \mathbf{z}_k - \mathbf{s}) \right\}, \quad (8)$$

where  $T_\beta\{\cdot\}$  is a soft-thresholding function which is applied to each element of its vectorial argument and is defined by

$$T_\beta\{y\} = \begin{cases} y(1 - \beta/|y|) & \text{if } |y| \geq \beta \\ 0 & \text{if } |y| < \beta \end{cases} \quad (9)$$

(b)

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}. \quad (10)$$

(c)

$$\mathbf{z}_{k+1} = \mathbf{x}_k + \frac{t_{k-1}}{t_{k+1}} (\mathbf{x}_k - \mathbf{x}_{k-1}). \quad (11)$$

(d) Check convergence or stopping condition.

Step 1 is required to prevent the argument of the soft-thresholding function to become negative. To find the maximum eigenvalue we use the Rayleigh's power method, which is an efficient technique to find an approximation of the dominant eigenvalue of a matrix, e.g. see [Larson and Edwards \(1999\)](#).

### 3 NUMERICAL EXAMPLES

#### 3.1 First example: synthetic data

We generated a synthetic NMO-corrected gather consisting of 13 traces with  $\theta_i \in (0^\circ, 36^\circ)$  and 13 reflectors with times between 0.0 and 1.6 s. The AVA response was modeled using the two-term Shuey approximation, then equation (2) becomes

$$r_t(\theta) = I_t + G_t \sin^2(\theta). \quad (12)$$

The coefficients  $I_t$  and  $G_t$ , which are the objective of the inversion and are functions of the physical properties of the media at each side of an interface, are known as *Intercept* and *Gradient* ([Shuey, 1985](#)). To test the method against noisy data we added Gaussian noise with standard deviation  $\sigma = \max_{t_i} |s_t(\theta_i)|/\text{SNR}$ , where SNR is the signal-to-noise ratio. In this particular example we set SNR=10. The data was generated using a Ricker wavelet ([Ricker, 1940](#)) with central frequency  $f_0 = 30$  Hz. Figures 1(a) and (b) show the *Intercept* and *Gradient* values used to generate the data and Figures 1(c) and (d) show the noise-free data and the noisy data.

Before applying FISTA to minimize equation (7) we need to select an appropriate trade-off parameter. There are various methods to estimate  $\lambda$ , such as the L-curve criterion, the discrepancy principle, and the generalized cross-validation criterion ([Farquharson and Oldenburg, 2004](#)). For the numerical examples we decided to estimate the trade-off parameter using the

discrepancy principle. This formulation is preferred when an estimation of the noise level is available. In practice, we construct the  $l_1$ -curve of the solution, also known as Pareto curve (Berg and Friedlander, 2009), and chose  $\lambda$  as the one that minimizes the  $l_1$ -norm while the  $l_2$ -norm of the errors (i.e. the misfit) remains less or equal than the noise level.

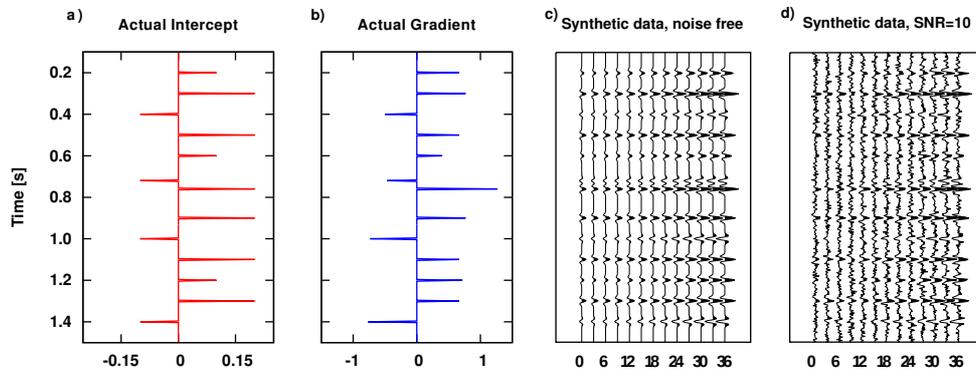


Figure 1: Actual *Intercept* (a) and *Gradient* (b) used to generate the synthetic data. c) Noise-free synthetic data. d) Noisy data with SNR=10.

Figure 2 shows the  $l_1$ -curves corresponding to the noise-free and the noisy gathers. Clearly, the higher the  $\lambda$ , the lower the  $l_1$ -norm, and *vice versa*. Through the discrepancy principle we estimated a value of  $\lambda = 0.25$  for the data with SNR = 10.

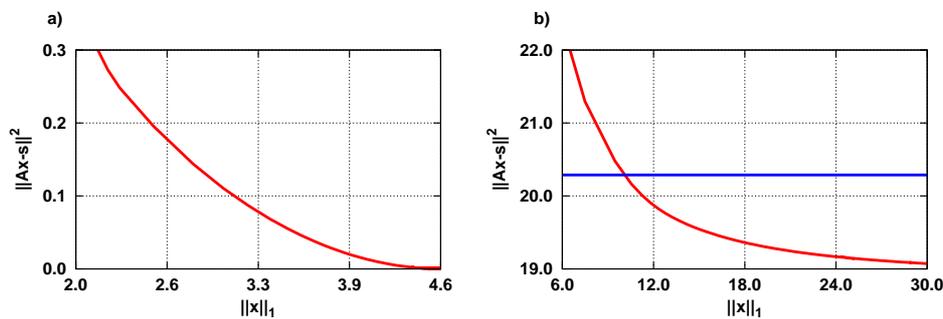


Figure 2: Red line:  $l_1$ -curve corresponding to a) noise-free data, b) noisy data with SNR=10. The blue line indicates the noise level.

Figure 3(a) shows the actual *Intercept* used to generate the data, Figure 3(b) shows the *Intercept* estimated from the noise-free data, and Figure 3(c) shows the *Intercept* estimated from the noisy data. In the second case (noisy data), we can observe that the results are very accurate even in the presence of noise. The solution shows some spurious spikes, but their amplitudes are relatively small. Figures 3(d) and 3(e) show the results of the inversion of the noisy data using conventional methods such as the least-squares (LS) inversion or the *Intercept/Gradient* sample-by-sample analysis (Yilmaz, 2001), respectively. The low resolution of the solutions obtained by these two techniques makes it difficult to interpret the estimated attributes. In the case of the conventional LS inversion, a prewhitening was used to stabilize the inversion, thus the amplitudes of the attributes are underestimated. The results show that under noisy conditions the solutions obtained by FISTA are much more accurate than the ones obtained with a conventional strategy.

Figure 3(f) shows the actual *Gradient* used to generate the data while Figures 3(g) to 3(j) show the estimated *Gradient* using FISTA and the conventional methods. As in the case of the

*Intercept*, in the presence of noise FISTA estimated the *Gradient* much more accurately than the conventional strategies.

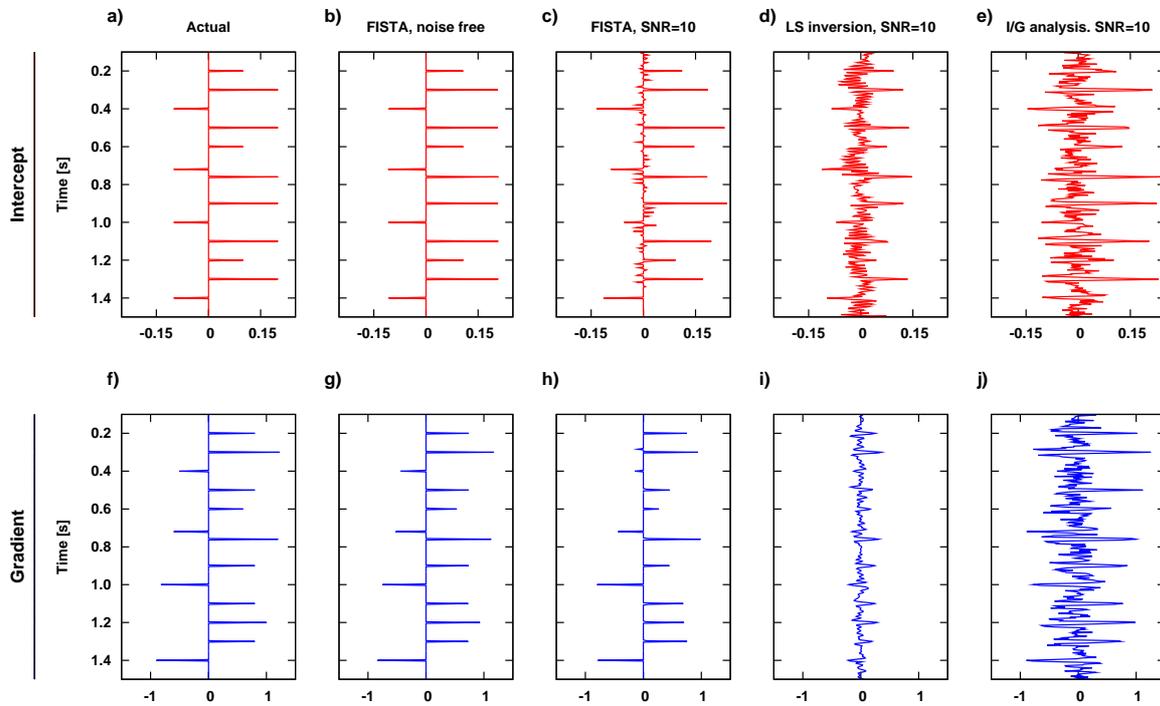


Figure 3: Actual *Intercept* (a) and *Gradient* (f). Estimated *Intercept* and *Gradient*: b) and g) from noise-free data, c) and h) from data with SNR=10, d) and i) using the conventional LS, e) and j) using the *Intercept/Gradient* analysis.

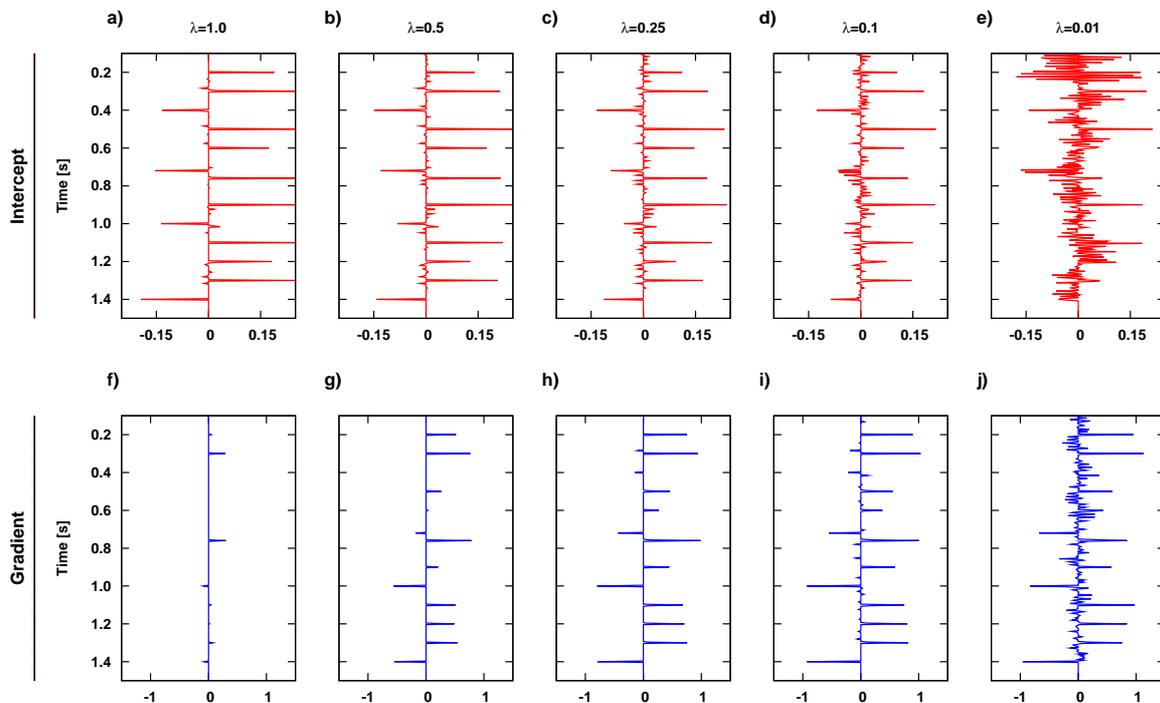


Figure 4: Estimated *Intercept* and *Gradient* from data with SNR=10 using a) and f)  $\lambda = 1.0$ , b) and g)  $\lambda = 0.5$ , c) and h)  $\lambda = 0.25$ , d) and i)  $\lambda = 0.1$ , e) and j)  $\lambda = 0.01$ .

Figure 4 shows that when  $\lambda$  is selected smaller than the value given by the discrepancy principle, which is equal to 0.25, the amplitudes of the spurious spikes both in the *Intercept* and the *Gradient* might become larger than acceptable. On the other hand, if  $\lambda$  is larger than this value, the estimated *Gradient* might be severely underestimated. The last issue is expected because the *Gradient* has a larger scale than the *Intercept*. Thus, for  $\lambda$  larger than optimal value, the  $l_1$ -norm term in equation (7) becomes significantly more sensitive to the *Gradient* variations than the misfit term. Then, FISTA tends to reduce at each iteration the  $l_1$ -norm of the *Gradient* more than the  $l_1$ -norm of the *Intercept*, without affecting the misfit significantly.

### 3.2 Second example: field data

The field data consist of 110 NMO-corrected angle-gathers with a sampling interval of 4 ms and a time window from 0.1 s to 0.5 s. To improve the lateral continuity the inversion was done on “super-gathers”, where each super-gather was built by averaging five consecutive gathers. Once inverted, the array containing the estimated *Intercept* and *Gradient* coefficients was assigned to the center gather. The next super-gather adds the next gather and discards the first one, and so the inversion process is repeated until all gathers are processed. The wavelet used in the inversion was estimated from each super-gather by assuming zero-phase (Robinson and Treitel, 2002).

Figures 5(c) and 5(d) show the resulting high-resolution *Intercept* and *Gradient* images obtained using FISTA. All significant reflectors were resolved correctly and a good lateral continuity was obtained, showing the consistency of the proposed method. From the estimated AVA coefficients we reconstruct the gathers and calculate the corresponding stack which is shown in Figure 5(b). This is quite similar to the original stacked data (Figure 5(a)), showing that the inversion honors the original input data.

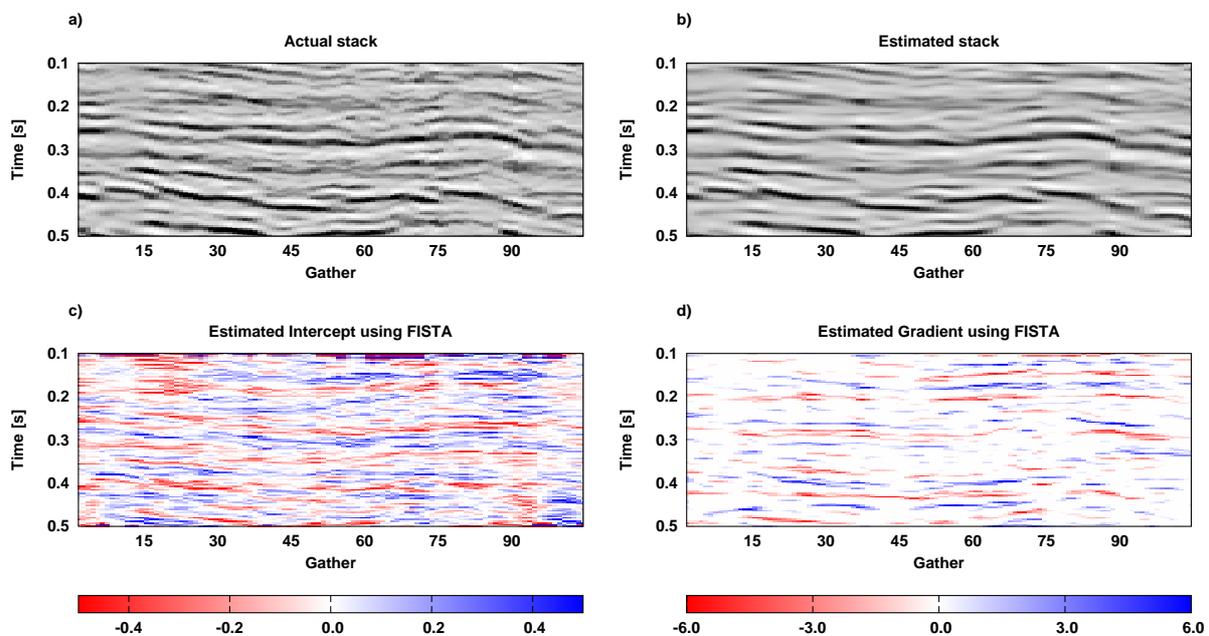


Figure 5: a) Actual stack, b) estimated stack from the inverted data, c) and d) high-resolution inverted *Intercept* and *Gradient* image using FISTA.

Figures 6(a) and 6(b) show the *Intercept* and *Gradient* images obtained using the conventional LS inversion. As in the case of the synthetic data example, the amplitudes of the estimated solutions were very much affected by the prewhitening used to stabilize the inversion, specially the *Gradient* image, that was significantly underestimated. Finally, Figures 6(c) and 6(d) show the *Intercept* and *Gradient* images obtained using the *Intercept/Gradient* sample-by-sample analysis. As can be seen, the resolution of the solutions obtained using the method proposed in this work is higher than the obtained using the conventional methods.

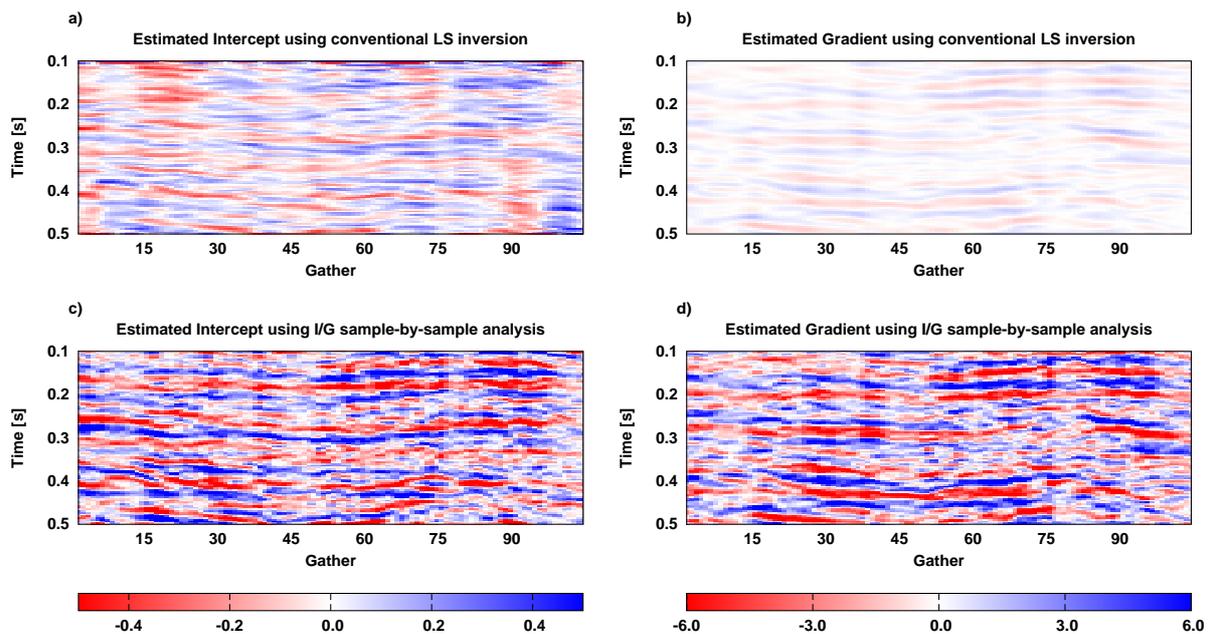


Figure 6: a) and b) Estimated *Intercept* and *Gradient* section using conventional LS inversion, c) and d) estimated *Intercept* and *Gradient* section using a *Intercept/Gradient* analysis.

## 4 CONCLUSIONS

In this work we presented a sparse-spike AVA inversion technique which is based on the Fast Iterative Shrinkage-Thresholding Algorithm. The method aims to obtain high-resolution AVA attributes from prestack data, such as *Intercept* and *Gradient*, minimizing both the misfit between the modeled and the observed data, and the  $l_1$ -norm of the solution. The numerical tests on synthetic data showed that the method is robust in presence of noise and provides very good results in all inverted magnitudes. Comparisons against conventional methods showed that the proposed strategy is capable to obtain solutions with higher resolution, and therefore easily to interpret. Because FISTA only uses matrix-vector multiplications, the method is economic in terms of computational cost. Numerical tests on field data showed that the proposed technique is capable of obtaining high-resolution *Intercept* and *Gradient* images that honor the observed data, improving the solutions obtained using conventional methods.

## 5 ACKNOWLEDGMENTS

We are grateful to Juan Carlos Soldo for giving us access to the field data set. This work was partially supported by Agencia Nacional de Promoción Científica y Tecnológica (PICT-2010-2129).

## REFERENCES

- Aki K. and Richards P. *Quantitative seismology: theory and methods*. W.H. Freeman and Co., 1980.
- Alemie W.M. and Sacchi M.D. High-resolution three-term AVO inversion by means of a trivariate Cauchy probability distribution. *Geophysics*, 76(3):R43–R55, 2011.
- Beck A. and Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal of Imaging Sciences*, 2:183–202, 2009.
- Berg E. and Friedlander M. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal of Scientific Computing*, 31(2):890–912, 2009.
- Castagna J.P., Swan H.W., and Foster D.J. Framework for AVO gradient and intercept interpretation. *Geophysics*, 63(3):948–956, 1998. ISSN 3. doi:10.1190/1.1444406.
- Daubechies I., Defrise M., and Mol C.D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57:1413–1457, 2004.
- Downton J.E. and Lines L.R. High-resolution AVO analysis before NMO. In *Expanded Abstracts*, volume 22, pages 219–222. SEG, 2003. ISSN 1. doi:10.1190/1.1817781.
- Farquharson C.G. and Oldenburg D.W. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophysical Journal International*, 156:411–425, 2004.
- Fatti J.L., Smith G.C., Vail P.J., Strauss P.J., and Levitt P.R. Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the Geostack technique. *Geophysics*, 59(9):1362–1376, 1994.
- Ikelle L.T. and Amundsen L. *Introduction to petroleum seismology*. Investigations in Geophysics. SEG, 2005.
- Larson R. and Edwards B.H. *Elementary Linear Algebra*. Houghton Mifflin Company, 4th edition, 1999.
- Misra S. and Sacchi M.D. Global optimization with model-space preconditioning: Application to AVO inversion. *Geophysics*, 73(5):R71–R82, 2008.
- Pérez D.O. and Velis D.R. Sparse-spike AVO/AVA attributes from prestack data. In *Expanded Abstracts*, volume 30, pages 340–344. SEG, 2011. ISSN 1. doi:10.1190/1.3627906.
- Ricker N. The form and nature of seismic waves and the structure of seismograms. *Geophysics*, 5:348–366, 1940.
- Robinson E.A. and Treitel S. *Geophysical Signal Analysis*. SEG, 2002.
- Shuey R. A simplification of the Zoeppritz equations. *Geophysics*, 50(4):609–614, 1985.
- Smith G.C. and Gidlow M. A comparison of the fluid factor with  $\lambda$  and  $\mu$  in AVO analysis. In *Expanded Abstracts*, pages 122–125. Soc. of Expl. Geophys., 2000.
- Yilmaz O. *Seismic Data Analysis: processing, inversion, and interpretation of seismic data*. Investigations in Geophysics. SEG, 2001.
- Zoeppritz K. Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeits-flächen. *Gott. Nachr. Math. Phys*, K1:66–84, 1919.