High-resolution prestack seismic inversion using a hybrid FISTA least-squares strategy

Daniel O. Pérez[1], Danilo R. Velis[1], and Mauricio D. Sacchi[2]

ABSTRACT

A new inversion method to estimate high-resolution amplitude-versus-angle attributes (AVA) attributes such as *intercept* and *gradient* from prestack data is presented. The proposed technique promotes sparse-spike reflectivities that, when convolved with the source wavelet, fit the observed data. The inversion is carried out using a hybrid two-step strategy that combines fast iterative shrinkage-thresholding algorithm (FISTA) and a standard least-squares (LS) inversion. FISTA, which can be viewed as an extension of the classical gradient algorithm, provides sparse solutions by minimizing the misfit between the modeled and the observed data, and the l_1-norm of the solution. FISTA is used to estimate the location in time of the main reflectors. Then, LS is used to retrieve the appropriate reflectivity amplitudes that honor the data. FISTA, like other iterative solvers for l_1-norm regularization, does not require matrices in explicit form, making it easy to apply, economic in computational terms, and adequate for solving large-scale problems. As a consequence, the FISTA+LS strategy represents a simple and cost-effective new procedure to solve the AVA inversion problem. Results on synthetic and field data show that the proposed hybrid method can obtain high-resolution AVA attributes from noisy observations, making it an interesting alternative to conventional methods.

INTRODUCTION

The analysis of the variations in amplitude of a reflected wave versus angle of incidence (AVA) or versus offset (AVO) plays a central role for hydrocarbon detection and lithology identification (Chopra and Castagna, 2007), for these variations depend on the contrasts between rock physical properties such as P-wave velocities, S-wave velocities, and densities. Given a plane wave that arrives at an interface that separates two adjacent media, AVA changes can be described by the Zoeppritz equations (Zoeppritz, 1919). Because these equations are nonlinear, they are impractical for applications such as data interpretation and inversion, and thus, various authors have developed several "linear" approximations over the last decades (e.g., Bortfeld, 1961; Aki and Richards, 1980; Shuey, 1985; Fatti et al., 1994). The coefficients of these approximations constitute AVA attributes that may provide important information about fluid content, a key issue for the characterization of hydrocarbon reservoirs (Ostrander, 1984; Castagna et al., 1998; Smith and Gidlow, 2000; Chopra and Castagna, 2007).

Given a set of prestack data, it is possible to devise an inverse strategy to retrieve the coefficients of these approximations, and ultimately, the model parameters that characterize the subsurface. However, because in general the inverse problem is ill posed, while there exist several sets of coefficients that honor the data equally well (nonuniqueness), some of these might exhibit a huge l_2-norm, thus being unstable and meaningless solutions. Reliable solutions can be obtained by regularization, that is, by promoting certain solutions through the minimization of an appropriate norm used to take into account useful prior information (Tarantola, 1987; Ulrych and Sacchi, 2005). For example, sparseness is a property that can be obtained using l_1-norm regularization, and that has been effectively applied to solve the seismic deconvolution problem by various authors (e.g., Taylor et al., 1979; Oldenburg et al., 1983; Sacchi, 1997). Sparse solutions are desirable because they lead to sharply resolved reflectors that overcome the band-limitation of classical l_2-norm solutions (Debeye and van Riel, 1990). Though we know from impedance well logs that the real earth structure is continuous rather than sparse-spike (Cooke and Schneider, 1983), it is also true that in general, the distribution of the amplitudes of reflection coefficients is nongaussian (Walden and Hosken, 1986; Velis, 2003). Moreover, analysis of well-bore-derived reflectivity sequences

indicate that main lithological units can be associated with sparsely distributed large reflection coefficients (Walden and Hosken, 1986). These facts have a profound significance on deconvolution, as demonstrated by the various approaches to deconvolution that rely on the non-Gaussianity assumption (e.g., Wiggins, 1978; Oldenburg et al., 1983; Robinson and Ebrom, 1996; Sacchi et al., 1996).

In the context of AVA inversion, methods that promote sparse-spike solutions have been studied with very interesting results by several authors. Some have proposed a Bayesian inversion where sparseness is obtained through the use of appropriate long-tailed prior probability distributions (Downton and Lines, 2003; Misra and Sacchi, 2008; Alemie and Sacchi, 2011). Alternatively, sparseness can be achieved by fixing the number of reflectors and using global optimization algorithms such as simulated annealing to obtain their locations, and least squares (LS) to estimate their amplitudes (Pérez and Velis, 2011). However, the sparse-spike high-resolution solutions provided by these techniques are achieved at the expense of an increase in the mathematical complexity and computational cost, important issues in current applications where data volumes are huge.

With this motivation, the method proposed, which shares the same objectives of techniques that honor the data (i.e., sparse-spike solutions), introduces a simple and cost-effective new procedure to solve the AVA inverse problem: the fast iterative shrinkage-thresholding algorithm (FISTA). FISTA is a powerful algorithm that has been presented very recently to minimize, together with the misfit term, the l_1-norm of the solution (Beck and Teboulle, 2009). 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. At each iteration, only matrix-vector multiplications, and no matrix inversions, are required. ISTA is known to have linear convergence (Bredies and Lorenz, 2008), but FISTA is shown to be faster by several orders of magnitude. In practice, the misfit and the l_1-norm are combined into a cost function by means of a trade-off parameter. In this context, FISTA provides sparse-spike solutions in a simple and effective way. However, in the case of AVA inversion of noisy data, the solutions might be correct from the mathematical point of view but physically meaningless, as it will be shown in the numerical examples. This is not an issue of FISTA alone, for it is just a tool for solving the optimization problem. The issue arises because the various parameters involved in the inversion (e.g., intercept and gradient when using Shuey's [1985] two-term approximation) may exhibit a different scale, and thus, their impact on the l_1-norm is not properly balanced. To alleviate this problem, we propose to use the solution obtained by FISTA as a priori information to perform, then in a second step, an LS inversion. This *debiasing* step (Figueiredo et al., 2007) leads to a hybrid technique that promotes sparse-spike solutions which are much more robust from the physical point of view than those obtained using l_1-*norm* regularization, with a negligible increase of the computational cost.

It is well known the inability of inversion methods to properly invert the gradient without having a priori knowledge of their response. Although intercept and gradient are sensitive to noise, gradient shows higher sensitivity (Herrmann and Cambois, 2001; Whitcombe et al., 2004). The variances of these two attributes are partially controlled by the geometry of the data acquisition (Downton et al., 2000). The standard-deviation of the gradient is larger than the standard-deviation of the intercept and increases with traveltime, making its conventional estimation inaccurate and its use questionable (Cambois, 1998). Moreover, the inclusion of a third-order term in the AVO analysis (e.g., curvature) may lead to even higher inaccuracies and instabilities because data often do not contain information to properly resolve the third-order term, except for wide-azimuth data and high signal-to-noise ratios (S/Ns) (Cambois, 2001). One way to alleviate the aforementioned problems is via the inclusion of a covariance matrix or scale matrix that provides a correlation between those parameters. This has been explored in the context of Gaussian statistics (Downton and Lines, 2001, 2004; Buland and Omre, 2003) and using sparse priors derived from a multivariate Cauchy distribution (Alemie and Sacchi, 2011). However, these methods assume that one knows the correlation or scale matrix that relate intercept and gradient behaviors. In this paper, we circumvent knowing the correlation matrix of the unknown parameters by using the proposed hybrid approach. Due to the type of angular aperture that we are using, we focus the analysis on the two-term Shuey's approximation, though the extension to higher-order approximations is straighforward. In our method, sparsity is used to estimate the position of the reflections by solving an underdetermined problem where it is difficult to produce stable and reliable estimates of the gradient. Then these traveltimes are used to simultaneously estimate intercept and gradient by solving a more stable, overdetermined LS problem. A key element of this paper is that we can stabilize the inversion of the gradient without having to include a covariance or scale matrix. This is important because, in general, the covariance matrix is only known when we have borehole measurements or a good understanding of the expected AVO classes in the area of study (Downton, 2005).

The paper is organized as follows. First, we explain the proposed method, setting up the necessary hypothesis and defining all relevant equations. In addition, we provide a step-by-step description of FISTA. Next, we test the method on normal-moveout (NMO) corrected prestack synthetic data assuming the classical two-term Shuey's approximation (Shuey, 1985) of the Zoeppritz equations. In the first numerical example, we show that, when the data are noisy, FISTA alone is not capable of providing satisfactory solutions to the intercept and gradient. Then, in a second numerical example, we show that the solutions obtained using the proposed hybrid method outperform the solutions obtained by FISTA, providing accurate estimates of intercept and gradient. Finally, we test the hybrid algorithm using a 2D field data set. We show that reliable, high-resolution intercept and gradient images that honor the observed data can be obtained.

THEORY

As usual, we rely on the convolutional model of the seismic trace and assume that the earth structure can be represented by a series of horizontal layers of constant material properties separated by planar interfaces (e.g., Yilmaz, 2001). Hence, an N-trace angle-gather can be expressed as

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

where $\mathbf{s}(\theta_i)$ is the seismic trace corresponding to the ith angle of incidence θ_i, $\mathbf{w}$ is the source wavelet of dimension L_w, $\mathbf{r}(\theta_i)$ is the reflectivity of dimension L_r, $\mathbf{n}(\theta_i)$ is the random noise, and * denotes convolution. Both $\mathbf{n}(\theta_i)$ and $\mathbf{s}(\theta_i)$ have dimension $L_s = L_w + L_r - 1$.

Assuming that the gather has been NMO corrected, and considering the validity of the well-known approximation to the Zoeppritz equations, the reflection coefficient as a function of incidence angle for a reflector at time t can be written, in a general form, as

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

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 $g_k(\theta_i)$ 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,

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

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

$$\mathbf{A}(\theta_i) = [\mathbf{A}_1(\theta_i)| \cdots |\mathbf{A}_n(\theta_i)]. \qquad (4)$$

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

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

for $h = 1, \cdots, L_s$ and $j = 1, \cdots, L_r$. Then, the N systems of equations given by expression 3 can be rearranged into the following unique system:

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

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 $\mathbf{s}(\theta_i)$, for $i = 1, \cdots, N$.

Equation 6 represents an overdetermined system of linear equations with nL_r unknowns, which are the coefficients of the selected Zoeppritz's approximation given by expression 2. As in the case of the deconvolution of reflection seismograms (Oldenburg et al., 1983), a stable and sparse solution can be obtained using an l_2-norm misfit with l_1-norm regularization. Thus, the problem reduces to find the $\mathbf{x}$ that minimizes the cost function

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

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. The second term is used to penalize nonsparse solutions. The trade-off parameter λ is used to balance the weight or impact of the two terms. Equation 7 can be minimized using FISTA (Beck and Teboulle, 2009).

At this point, it is important notice that, as mentioned in the introduction, the minimization of equation 7 under noisy conditions may lead to solutions which are mathematically correct but unstable. This is because the different coefficients involved in the second term of equation 7 may have different magnitudes, and thus their impact on the minimization of J would be different. This issue will be illustrated in the next section where we show that, for noisy data,

when λ is large, the solutions may honor the observed data but the amplitudes of some of the estimated coefficients are severely underestimated. However, for λ small, the solutions may honor the data but the amplitudes of some of the estimated coefficients are severely overestimated. For these reasons, we propose to carry out the AVA inversion by means of a two-stage procedure. In the first stage, we use FISTA to minimize J using an appropriate value for λ such that the solution is sufficiently sparse (in the following sections, we discuss how to select this trade-off parameter). The solution provided by the first stage is then used as a priori information that provides the "support" of the solution (i.e., the times associated with the non-zero reflectors) that is to be used for the second stage. The second stage consists of a LS optimization that optimizes the amplitudes of the unknowns by minimizing the misfit term only. We point out that the stability of the inversion is guaranteed provided the initial solution is sufficiently sparse. A similar two-stage strategy is proposed by Velis (2008) for solving the sparse-spike deconvolution, but the support was obtained, instead, using a much more computationally intensive algorithm such as simulated annealing.

FISTA stage

FISTA is an iterative algorithm devised to solve equations in the form of equation 7. Step-by-step, FISTA is summarized as follows:

1) Set $\alpha \geq \sigma_{max}$, where σ_{max} is 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 iteration $k = 1, 2, 3, \cdots$:
   a)

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

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

$$T_\beta\{g\} = \begin{cases} g(1 - \beta/|g|), & |g| \geq \beta, \\ 0. & |g| < \beta, \end{cases} \qquad (9)$$

   b)

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

   c)

$$\mathbf{z}_{k+1} = \mathbf{x}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_k - \mathbf{x}_{k-1}), \qquad (11)$$

   and
   d) check convergence or stopping condition.

Step 1 is required to prevent the argument of the soft-thresholding function from becoming negative. To find the maximum eigenvalue, we use the Rayleigh's power method (e.g., Larson and Edwards, 1999). For more details about FISTA, please refer to the work by Beck and Teboulle (2009).

LS stage

Once the support of the solution has been obtained using FISTA, we perform a debiasing step (Figueiredo et al., 2007) to adjust the

amplitudes of the coefficients. Now, given the M nonzero components at times τ_j of $\mathbf{x}$ (as output from FISTA, with $M \ll L_r$), their optimal amplitudes can be estimated using the classical LS criterion.

Following the same procedure to obtain equation 6, but assuming a sparse-spike model for the reflectivity, after some simple algebraic steps, it is easy to show that the system of equations that relates the data $\mathbf{s}$ with the unknown coefficients $\mathbf{y}$ becomes

$$\mathbf{By} = \mathbf{s}, \tag{12}$$

where $\mathbf{y}$ is the one-column vector formed by n blocks of dimension M containing the coefficients of the approximation given by equation 2, and $\mathbf{B}$ is an augmented matrix of dimension $L_s \times Mn$ formed by submatrices $\mathbf{B}_k$ with elements given by

$$\{\mathbf{B}_k(\theta_i)\}_{hj} = g_k(\theta_i) w_{h-\tau_j+1}, \tag{13}$$

for $h = 1, \ldots, L_s$, $j = 1, \ldots, M$, and $k = 1, \ldots, n$.

The solution to this new system of linear equations can be obtained by minimizing the squared differences between the observed and the modeled data, which leads to

$$\mathbf{y} = (\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{s}). \tag{14}$$

Essentially, FISTA is used to select the number M and the times τ_j of the spikes from the nonzero components of $\mathbf{x}$, while the LS inversion is used to estimate the values of their amplitudes.

In summary, the sparse-spike AVA-inversion procedure is as follows:

1) A sparse solution $\mathbf{x}$ of the equation 6 is estimated using FISTA.
2) The values of M and the τ_j, $j = 1, \cdots, M$, are determined by the nonzero components of $\mathbf{x}$.
3) Solution $\mathbf{y}$ is obtained by LS inversion (equation 14).

SYNTHETIC DATA NUMERICAL EXAMPLES

To illustrate the procedure, we generate a synthetic NMO-corrected gather consisting of 13 traces with $\theta_i \in (0^\circ, 36^\circ)$, and 12 reflectors in a time window of 0.5 s. The AVA response is modeled using the classical two-term Shuey's approximation, and so equation 2 becomes

$$r_t(\theta_i) = A_t + B_t \sin^2(\theta_i). \tag{15}$$

The coefficients A_t and B_t, which are the unknowns of the inversion, are known as intercept and gradient (Shuey, 1985). Figure 1a and 1b shows the intercept and gradient used to generate the data traces $s_t(\theta_i)$. The data were obtained by convolving $r_t(\theta_i)$ with a Ricker wavelet (Ricker, 1940) of central frequency $f_0 = 30$ Hz (Figure 1c). To test the method against noisy data we added Gaussian noise with standard deviation $\sigma = \max_{t,i} |s_t(\theta_i)| / $ S/N, where S/N is the signal-to-noise ratio (Figure 1d–1f).

First example: AVA inversion using FISTA only

To analyze the behavior of FISTA, in this first example, we perform the AVA inversion of the noisy data shown in Figure 1 using FISTA only. That is, we omit the LS stage described in the previous section. For this purpose, we first need to select an appropriate trade-off parameter λ in equation 7. In general, the selection of this parameter depends on the noise level of the data at hand. Clearly, if λ is too large, a very sparse solution would be obtained, but the data may not be properly fit. Conversely, if λ is too small, the noisy data would be overfit and the solution becomes unstable. Therefore, an appropriate strategy is required to select this key parameter. In this sense, there are various methods reported in the literature that can be used to estimate λ. For instance, one can adopt the so-called L-curve criterion, the discrepancy principle, or the generalized crossvalidation criterion (Farquharson and Oldenburg, 2004). For this numerical example, we estimate the trade-off parameter using the discrepancy principle. This formulation is preferred when an estimation of the noise level in the data is available. To this end, we construct the L-curve of the data, also known as *Pareto curve*
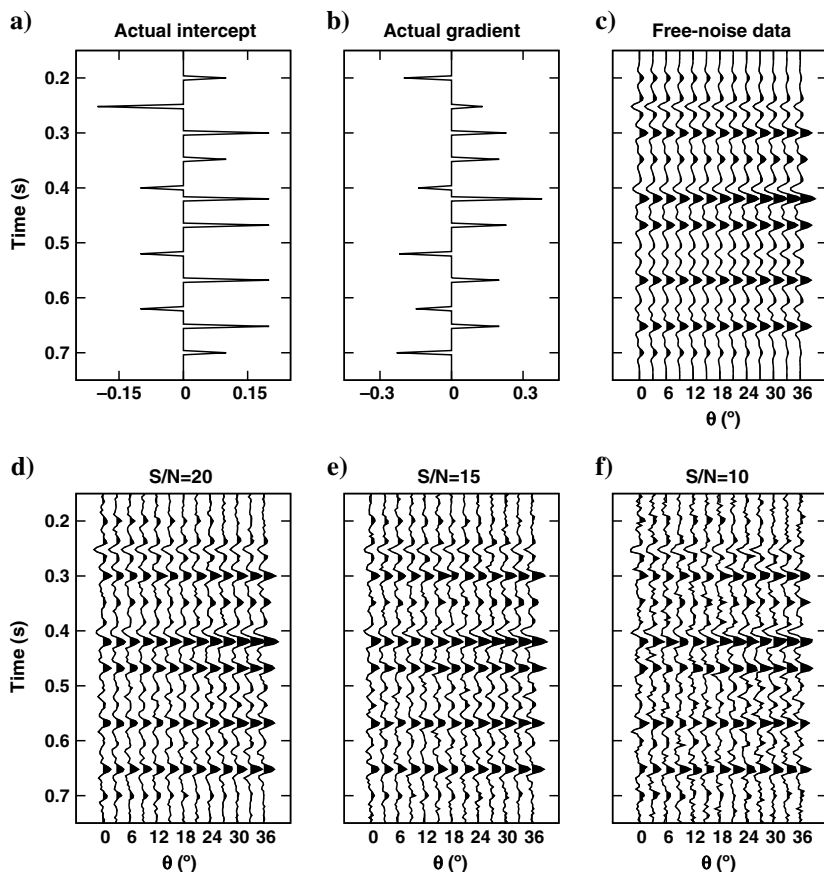

Figure 1. (a) Actual intercept and (b) actual gradient used to generate (c) the noise-free synthetic data, and the noisy data with (d) S/N = 20, (e) S/N = 15, (f) S/N = 10.

([Hennenfent et al., 2008](); [van den Berg and Friedlander, 2009]()), by carrying out the inversion using various trial λs and plotting the resulting l_1-norm versus misfit values. Then, we choose the optimum λ 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 2 shows the Pareto curves using FISTA only (long dashed lines) corresponding to the three noisy gathers shown in Figure 1. The horizontal solid lines indicate the noise level, and the small arrows show the direction of increasing λ. As expected, the larger the λ, the smaller the l_1-norm and the larger the misfit, and vice versa. Following the discrepancy principle, we estimated the optimum values $\lambda = 0.06$ for the data with S/N = 20, $\lambda = 0.1$ for the data with S/N = 15, and $\lambda = 0.14$ for the data with S/N = 10. These values correspond to the crossing of the Pareto and noise level curves, and are indicated with white circles in Figure 2.

Figure 3 shows the actual intercept and gradient, and the estimated attributes corresponding to the noise-free and noisy gathers. We can observe that the results obtained for the intercept are accurate in the presence of a significant amount of noise. The location in time of all the reflectors and their respective amplitudes were estimated correctly. Although the solutions show some spurious spikes, especially for the lowest S/Ns, their amplitudes are very small and they could be easily deleted by means of a threshold filter for an almost perfect match with the actual intercept. On the other hand, the solutions estimated for the gradient do not resemble the true gradient attribute used to model the data. The locations in time of the reflectors given by FISTA are correct, but their amplitudes are not. As the noise level increases, the amplitudes of the inverted spikes are severely underestimated. This is expected because the noise tends to mask the AVA response but not the energy of the reflector, and the misfit term in equation 7 becomes less sensitive to variations in the gradient than the l_1-norm term. Therefore, to decrease J as much as possible, FISTA tends to reduce the l_1-norm of the gradient more than the l_1-norm of the intercept, without affecting the misfit significantly. In cases of data with large amounts of noise (e.g., Figure 1f) where the AVA response is almost indistinguishable (such as those reflectors that exhibit a small gradient in Figure 3g), FISTA tends to reduce, for a given value of λ, the amplitude of the estimated gradient to zero. This is the case of the reflector located at $t = 0.25$ s, for example, which exhibits a small gradient but a significant intercept, and therefore, its gradient is annihilated after the inversion while its intercept is correctly estimated.

To continue the analysis of the behavior of the AVA inversion when using FISTA only, Figure 4 shows the results of the inversion for various values of λ in the case of the noisy gather with S/N = 20 (Figure 1d). When λ is smaller than the value suggested by the discrepancy principle, which is equal to 0.06, the amplitudes of the spurious spikes in the intercept and the gradient might become larger than acceptable. On the other hand, if λ is larger than this value, the estimated gradient might be severely underestimated, as shown in Figure 4i and 4j. This behavior is not observed in the estimated intercept because their scales are different, as explained in the previous paragraph. In this case, there is a wide range of λ values for which the estimated intercept amplitudes are acceptable. We conclude that we cannot find a value of λ leading to a correct estimation of intercept and gradient simultaneously, except for data with a high S/N, as shown in Figure 3.

Second example: AVA inversion using FISTA and LS

We have shown that FISTA alone does not give appropriate sparse solutions for the gradient, however the intercept can be estimated with a high degree of accuracy. The method proposed in this work takes advantage of the aforementioned result. In fact, we use the estimated intercept as a priori information to improve the estimation of the gradient through an LS inversion. We tested the proposed FISTA hybrid method (FISTA+LS) using the synthetic data described in the preceding section. Figure 2 shows the Pareto curves for FISTA+LS (short dashed lines) corresponding to the three noisy gathers portrayed in Figure 1. Compared to the Pareto curves associated with FISTA alone (long dashed lines), the curves associated with the hybrid method exhibit a significant break when they approach the noise level. Hence, when an estimation of the noise level is not available, it seems easier to estimate a value of λ using FISTA+LS than using FISTA only. Furthermore, Figure 5 (long dashed lines) shows the number of the nonzero spikes obtained after the inversion as a function of λ. Note that, for a given gather, there
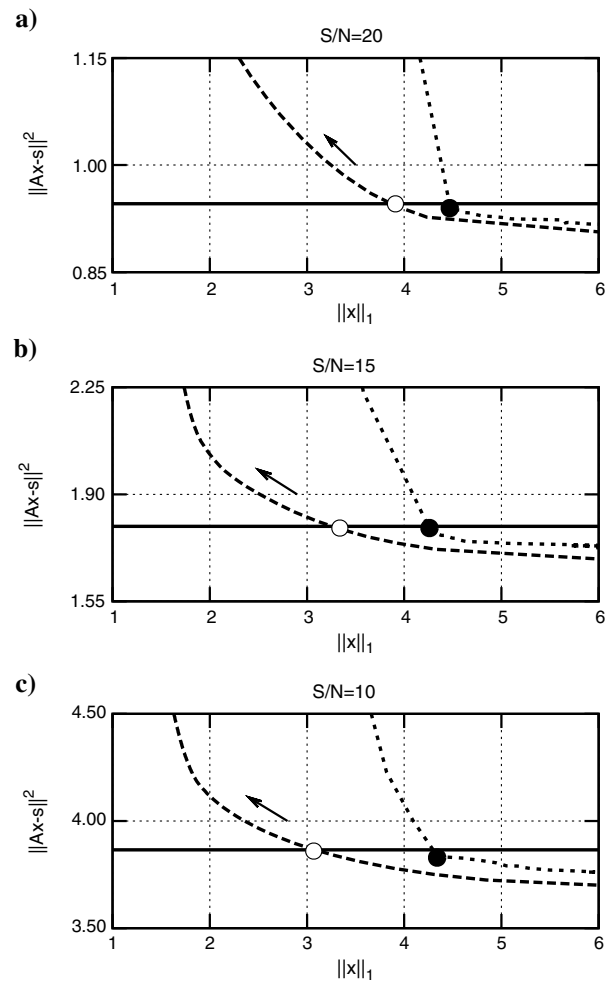

Figure 2. Pareto curves after the inversion of the noisy gathers shown in Figure 1 using FISTA only (long dashed lines) and FISTA+LS (short dashed lines), for various trial values of λ (the small arrows show the direction of increasing λ). The horizontal solid lines indicate the noise level, whereas white and black circles denote the selected optimum λ according to the discrepancy principle.
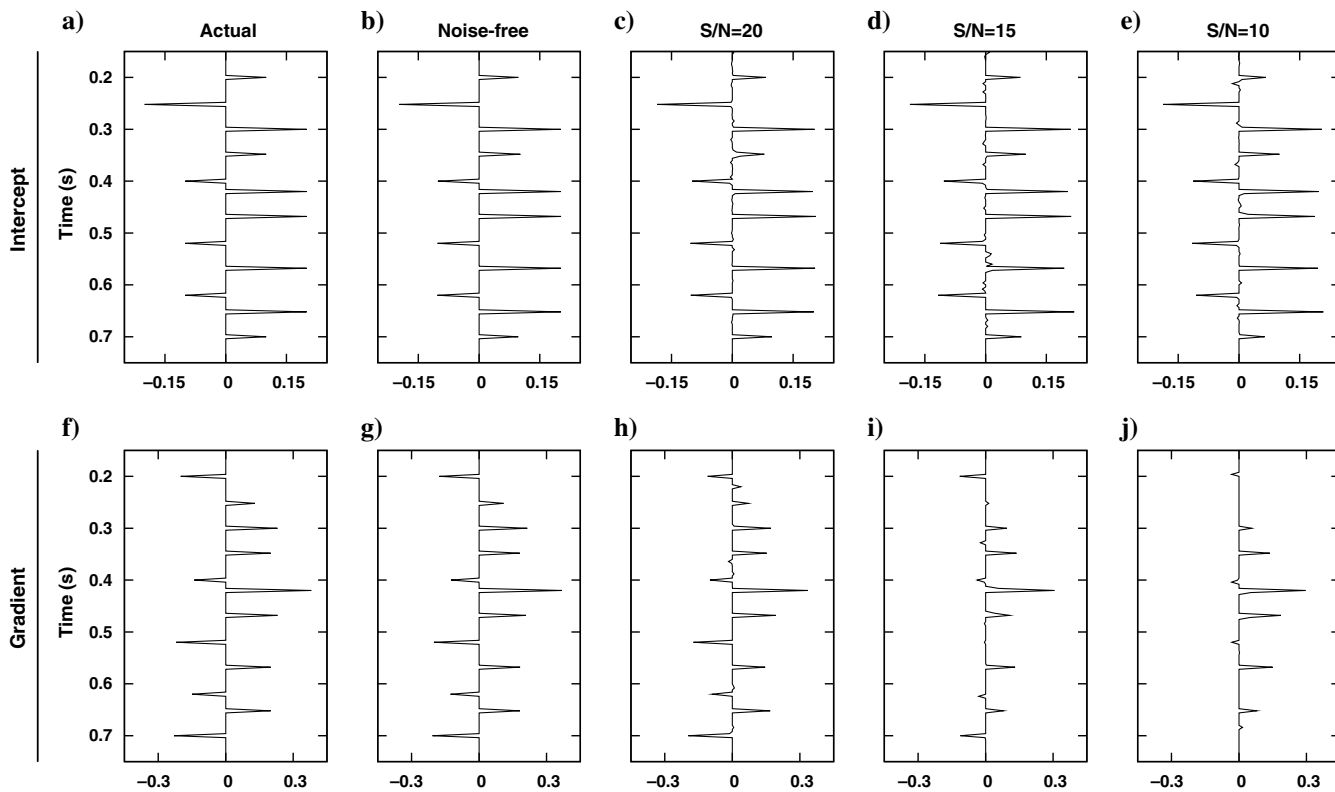
Figure 3. Results of the inversion using FISTA only: (a and f) Actual intercept and gradient, (b and g) estimated attributes corresponding to the noise-free data, (c and h) noisy data with S/N = 20, (d and i) S/N = 15, (e and j) S/N = 10.
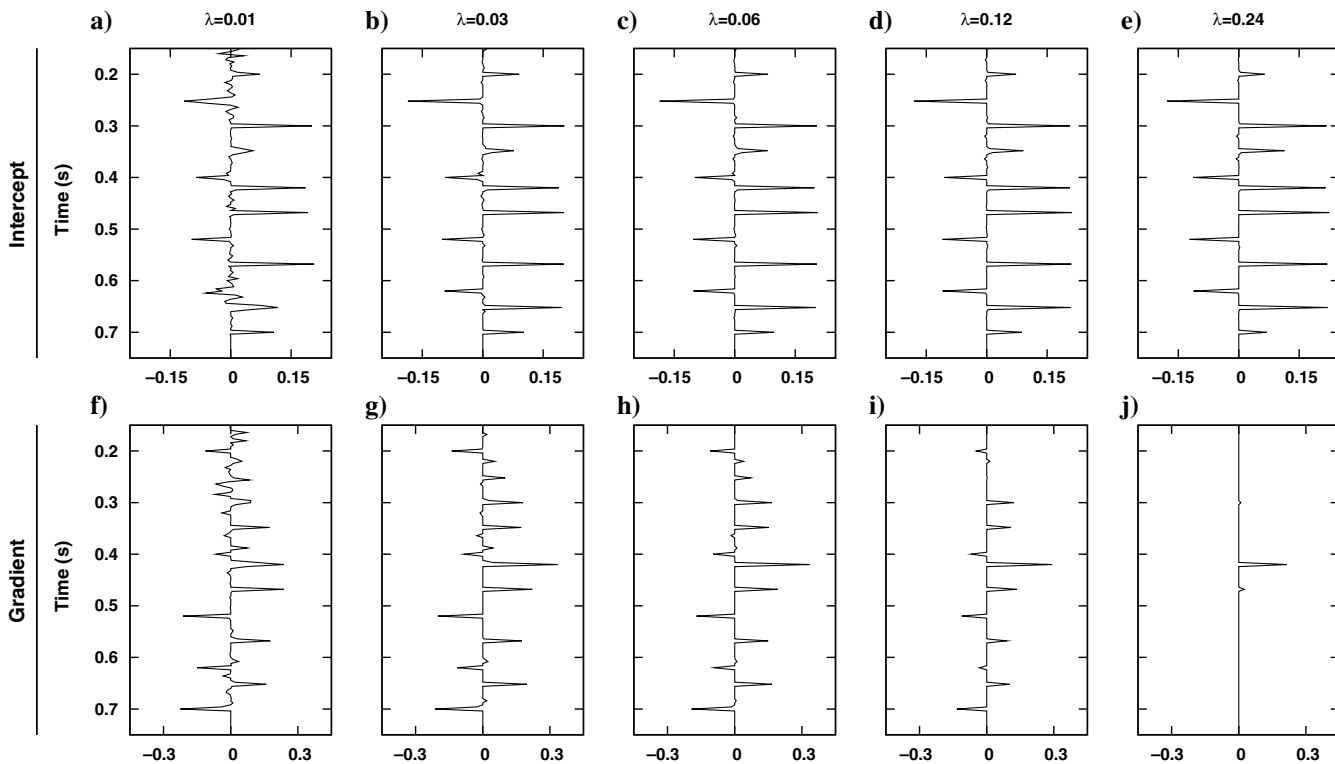

Figure 4. Results of the inversion using FISTA only for various values of λ in the case of the noisy gather with S/N = 20 (Figure 1d): (a and f) $\lambda = 0.01$, (b and g) $\lambda = 0.03$, (c and h) $\lambda = 0.06$, (d and i) $\lambda = 0.12$, (e and j) $\lambda = 0.24$.

is a relatively wide range of λ values for which the inversion yielded 12 reflectors, which is equal to the actual number of nonzero spikes of the model. These ranges are represented by the solid part of the curves shown in Figure 5, which map into the black circles shown in Figure 2. As a result, these black circles represent, in practice, a range of λ values instead of a single λ. Therefore, unlike FISTA alone, where according to the discrepancy principle a single "optimal" λ exists (white circles in Figure 5), when FISTA+LS is used there are several λ values that minimize the l_1-norm while the misfit remains less or equal to the noise level, a fact that is shown by inspecting the misfit as a function of λ in Figure 5 (short dashed lines). This means that the FISTA+LS solution depends, essentially, on the number of spikes provided by FISTA alone.

In summary, when FISTA is used without the LS stage, and because λ balances the impact of the two terms in the cost function (equation 7), the amplitudes of the estimated spikes are expected to decrease as λ increases, even when the number of nonzero spikes remains constant (as seen in Figure 4). But, on the other hand, when FISTA+LS is used, then the misfit and the amplitudes of the spikes after the LS stage are adjusted to the same values, irrespective of the selected λ within the range where the number of spikes remains constant (solid part of the curves in Figure 5).

Following these results, for this particular example we decided to use $\lambda = 4$ for the inversion of all the noisy gathers, though any other value within the range denoted by the solid part of the curves in Figure 5 would lead to exactly the same results.

Figure 6 shows the actual intercept and gradient, and the estimated attributes corresponding to the noise-free and noisy gathers. We can observe that FISTA+LS gives, in all cases, very accurate estimates for the intercept and the gradient. All reflectors and their amplitudes were adequately estimated, even under noisy conditions.

Finally, Figure 7 shows, for the case of the noisy-data with $S/N = 10$, a comparison among the solutions obtained using FISTA only, FISTA+LS, and those using two "conventional" methods. The conventional approaches include (1) a sample-by-sample intercept/gradient analysis and (2) a standard nonsparse LS inversion (Yilmaz, 2001). The first approach corresponds to the intercept and gradient that are obtained after a linear regression of the AVA amplitudes at each time sample. On the other hand, the standard LS inversion corresponds to the minimization of J with $\lambda = 0$, that is, by assuming a reflector at each time sample. Clearly, and because sparsity is not promoted, the low resolution of the attributes obtained by the conventional methods hampers the interpretation of the results. In the case of the conventional LS inversion, a moderate prewhitening was needed to stabilize the inversion. As a consequence, although data are honored very accurately (recall that $\lambda = 0$ in equation 7), the amplitudes of the derived attributes are severely underestimated. In the case of the conventional intercept/gradient analysis, amplitudes are not underestimated, but the solutions are unstable, especially the gradient. As already discussed, FISTA only can estimate an accurate intercept, however, the gradient is underestimated. In contrast, the results obtained by FISTA+LS are much more accurate than those obtained by all the other strategies, allowing one to retrieve stable sparse-spike like solutions for both attributes even under noisy conditions.

FIELD DATA NUMERICAL EXAMPLE

In this section, we test the method using field data. Field data complicates the application of our method because not only are

the noise level and the number of reflectors not known, but also the source wavelet is not available. We assume that the data have been properly processed to preserve amplitudes, and that the residual wavelet after source processing is zero phase. The field data consist of 110 NMO-corrected angle-gathers with a sampling interval of 4 ms. The inversion is carried out in a time window from 0.1 to 0.5 s. Rather than processing one gather at a time, and to improve the lateral continuity of the results, the inversion is applied to "supergathers," where each supergather was built by a weighted average of five consecutive gathers. Once inverted, the array containing the estimated intercept and gradient coefficients is assigned to the center gather.

The wavelet used in the inversion is estimated from each supergather by assuming zero phase (Robinson and Treitel, 2002).

Figure 8 shows the full stack section (7°–23°) of the observed seismic data (a), the full stack section reconstructed from the estimated solutions given FISTA+LS (b), the resulting high-resolution intercept and gradient images estimated using FISTA+LS (c and d) and FISTA only (e and f), and the low-resolution images obtained using the conventional methods (g to j). In addition, Figure 9 shows a selected input NMO-corrected gather (gather #45) and the corresponding reconstructions. We can observe that the reconstructed data are quite similar to the actual input data, showing that the
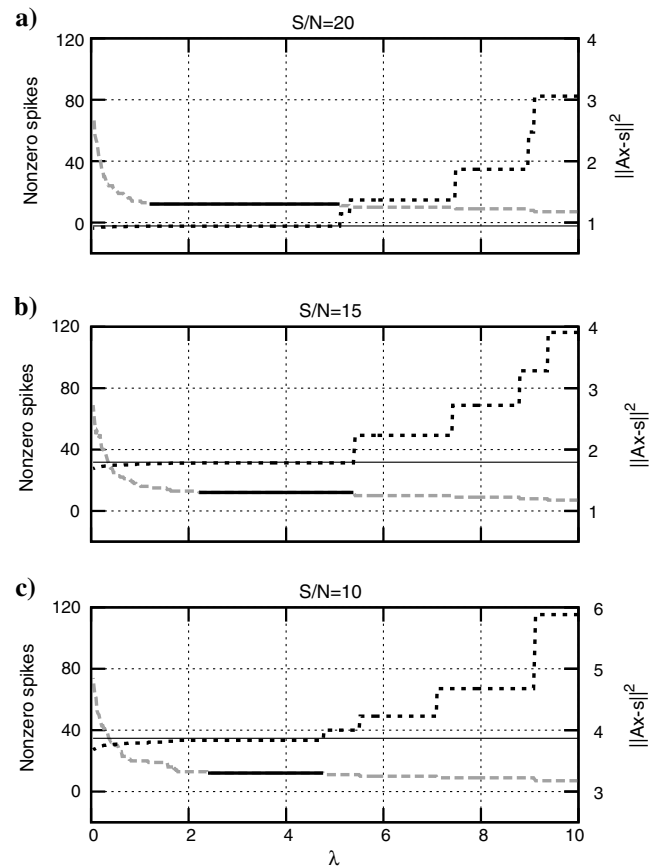

Figure 5. Number of nonzero spikes (long dashed lines) and misfit (short dashed lines) as a function of λ for the noisy data using FISTA +LS: (a) S/N = 20, (b) S/N = 15, (c) S/N = 10. The solid part of the curves that represent the number of nonzero spikes indicates the range of λ values for which the number of nonzero spikes is equal to the actual number of spikes of the model, which is 12. The horizontal thin solid lines indicate the noise level.
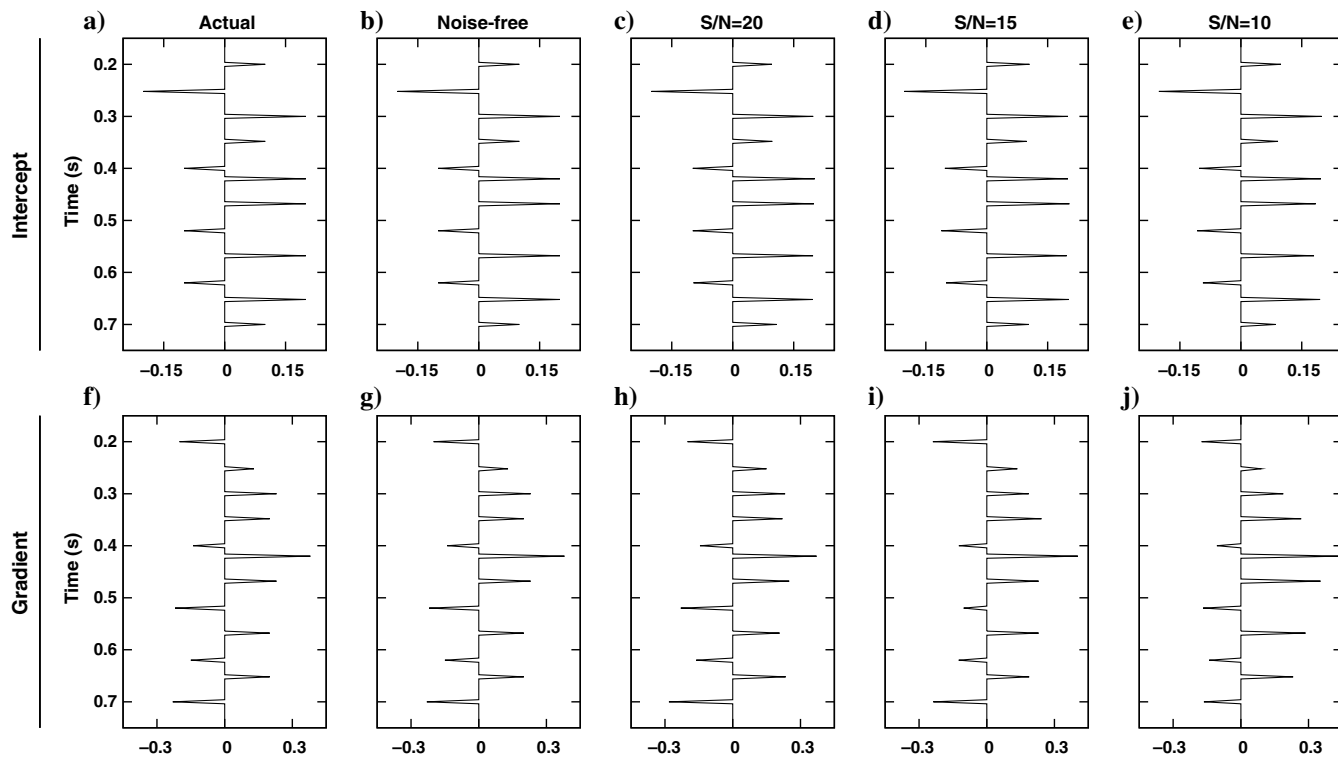
Figure 6. Results of the inversion using FISTA+LS: (a and f) Actual intercept and gradient, (b and g) estimated attributes corresponding to the noise-free data, (c and h) noisy data with S/N = 20, (d and i) S/N = 15, (e and j) S/N = 10.
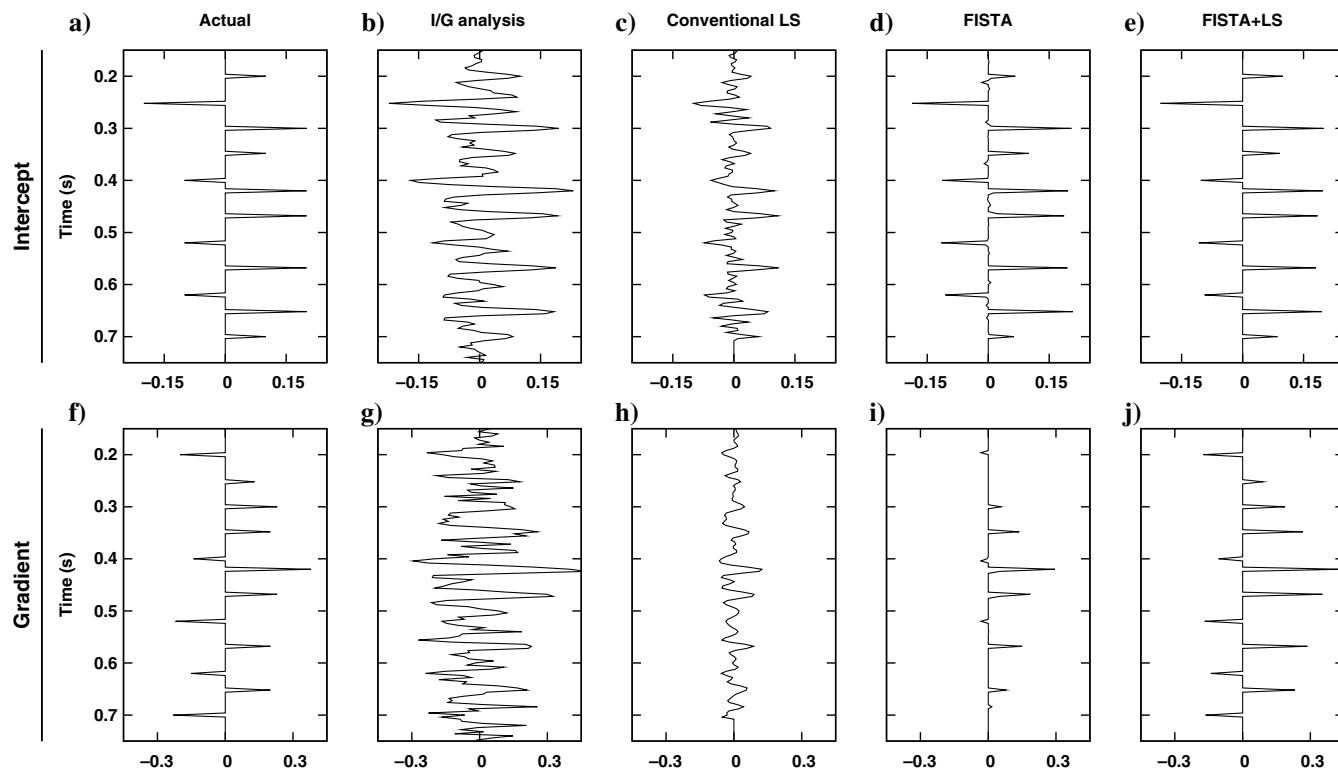
Figure 7. Comparison of the results of the inversion using various methods for the noisy data (S/N = 10). (a and f) Actual intercept and gradient. Estimated attributes using: (b and g) sample-by-sample intercept/gradient analysis, (c and h) conventional LS inversion, (d and i) FISTA only, and (e and j) FISTA+LS.
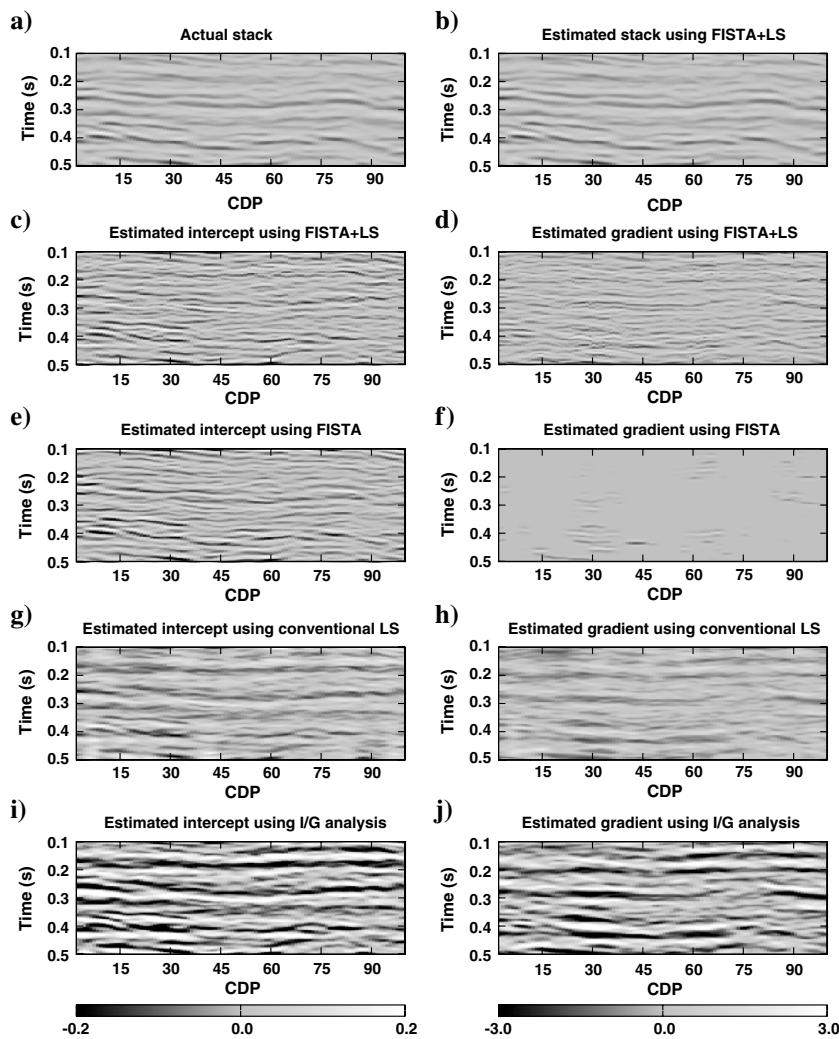
Figure 8. Field data example: (a and b) Actual and estimated stacks using FISTA+LS. Estimated attribute images using: (c and d) FISTA+LS, (e and f) FISTA only, (g and h) conventional LS inversion, (i and j) intercept/gradient analysis.
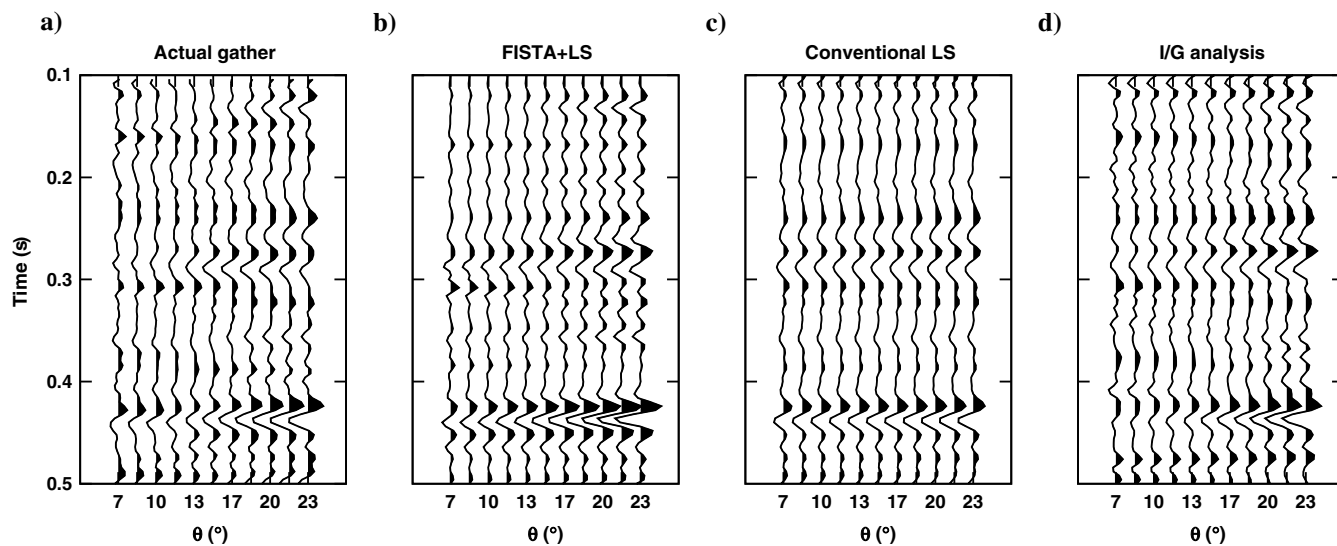
Figure 9. Field data example: (a) Actual NMO-corrected gather (gather 45), and the corresponding reconstructed gathers using (b) FISTA+LS, (c) conventional LS inversion, and (d) intercept/gradient analysis.

intercept and gradient inversions using FISTA+LS honor the observations very well. Contrary to the other approaches, FISTA+LS provides reasonable high-resolution attribute images for the intercept and gradient attributes, where all significant reflectors were resolved correctly.

As can be seen, there is a considerable difference in the magnitudes of the estimated intercept and gradient. According to Cambois (1998, 2001), the reason for this behavior is the limited angle range (from 7° to 23° in this example), and the presence of noise. Then, it is not unusual to observe differences of this order of magnitude when working on real data. Other authors attribute this difference to physical processes such as absorption, transmission losses, and anisotropy (Smith and Gidlow, 1987), or inadequate data processing before AVO inversion (e.g., incorrect NMO correction and inaccuracies in the estimated incidence angles), as suggested by Castagna and Smith (1994) and Jin et al. (2000). We stress that the method does not require a covariance or scale matrix as in the method proposed by Alemie and Sacchi (2011). As expected, FISTA correctly resolves the intercept; however, the amplitudes of the gradient are underestimated, as in the case of the synthetic data examples. As can be seen, the resolution and accuracy of the solutions obtained using the proposed method is higher than those obtained using the conventional methods. In particular, the images obtained by the conventional LS inversion are very much affected by the prewhitening used to stabilize the inversion. The latter is particularly evident in the gradient that was significantly underestimated. The intercept/gradient analysis does not show this problem, but the corresponding images clearly lack resolution.

CONCLUSIONS

We present a sparse-spike inversion method to obtain high-resolution AVA attribute from prestack data based on FISTA and LS inversion. The method aims to simultaneously obtain the location in time of the reflectors and the values of AVA attributes such as the intercept and gradient. As shown in the numerical examples, FISTA alone is not able to obtain an adequate gradient solution, because it strongly depends on the noise level and the chosen trade-off parameter λ. The use of the LS inversion as a debiasing step proved to be an appropriate choice to improve the solutions given by FISTA and obtain high-resolution AVA attributes that honor the observed data. It is worth noting that the proposed method allows us to stabilize the gradient solution without the use of a priori information or the use of a covariance or scale matrix. It is also remarkable, and different from other known methods, that there is no need to start the inversion process from an initial model close to the optimal solution. Here, the selection of λ is not a issue because a relatively wide range of λ values lead to similar solutions, and because FISTA+LS is much less sensitivity to noise level than FISTA alone. On noisy synthetic data, the proposed two-step hybrid strategy proves to be robust and stable, providing very accurate sparse-spike solutions for all the inverted magnitudes. Further, the solutions provided by FISTA+LS exhibit a much higher resolution than those given by two conventional methods. Tests on field data show that, contrary to the conventional methods, the proposed technique can provide high-resolution intercept and gradient images that accurately honor the observed data. Finally, because FISTA only uses matrix-vector multiplications and the matrices which need to be inverted to make the LS inversion are relatively small, the hybrid FISTA+LS method is economic in terms of computational cost.

REFERENCES

Aki, K., and P. Richards, 1980, Quantitative seismology: Theory and methods: W. H. Freeman and Co.

Alemie, W. M., and M. D. Sacchi, 2011, High-resolution three-term AVO inversion by means of a trivariate Cauchy probability distribution: Geophysics, **76**, no. 3, R43–R55, doi: 10.1190/1.3554627.

Beck, A., and M. Teboulle, 2009, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: SIAM Journal on Imaging Sciences, **2**, 183–202, doi: 10.1137/080716542.

Bortfeld, R., 1961, Approximations to the reflection and transmission coefficients of plane longitudinal and transverse waves: Geophysical Prospecting, **9**, 485–502, doi: 10.1111/j.1365-2478.1961.tb01670.x.

Bredies, K., and D. A. Lorenz, 2008, Linear convergence of iterative soft-thresholding: Journal of Fourier Analysis and Applications, **14**, 813–837, doi: 10.1007/s00041-008-9041-1.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185–198, doi: 10.1190/1.1543206.

Cambois, G., 1998, AVO attributes and noise: Pitfalls of crossplotting: 68th Annual International Meeting, SEG, Expanded Abstracts, 2027–2075.

Cambois, G., 2001, AVO processing: Myths and reality: 71st Annual International Meeting, SEG, Expanded Abstracts, 189–192.

Castagna, J. P., and S. W. Smith, 1994, Comparison of AVO indicators: A modeling study: Geophysics, **59**, 1849–1855, doi: 10.1190/1.1443572.

Castagna, J. P., H. W. Swan, and D. J. Foster, 1998, Framework for AVO gradient and intercept interpretation: Geophysics, **63**, 948–956, doi: 10.1190/1.1444406.

Chopra, S., and J. Castagna, 2007, Introduction to this special section — AVO: The Leading Edge, **26**, 1506–1507, doi: 10.1190/1.2821935.

Cooke, D., and W. Schneider, 1983, Generalized linear inversion of reflection seismic data: Geophysics, **48**, 665–676, doi: 10.1190/1.1441497.

Daubechies, I., M. Defrise, and C. D. Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: Communications on Pure and Applied Mathematics, **57**, 1413–1457, doi: 10.1002/cpa.20042.

Debeye, H. W. J., and P. van Riel, 1990, L_p-norm deconvolution: Geophysical Prospecting, **38**, 381–403, doi: 10.1111/j.1365-2478.1990.tb01852.x.

Downton, J., L. Lines, B. Goodway, Y. Xu, and Y. Li, 2000, Predicting the statistical properties of the fluid stack: 70th Annual International Meeting, SEG, Expanded Abstracts, 224–227.

Downton, J. E., 2005, Seismic parameter estimation from AVO inversion: Ph.D. thesis, University of Calgary.

Downton, J. E., and L. R. Lines, 2001, Constrained three parameter AVO inversion and uncertainty analysis: 71st Annual International Meeting, SEG, Expanded Abstracts, 251–254.

Downton, J. E., and L. R. Lines, 2003, High-resolution AVO analysis before NMO: 73rd Annual International Meeting, SEG, Expanded Abstracts, 219–222.

Downton, J. E., and L. R. Lines, 2004, Three term AVO waveform inversion: 74th Annual International Meeting, SEG, Expanded Abstracts, **23**, 215–218.

Farquharson, C. G., and D. W. Oldenburg, 2004, A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems: Geophysical Journal International, **156**, 411–425, doi: 10.1111/j.1365-246X.2004.02190.x.

Fatti, J. L., G. C. Smith, P. J. Vail, P. J. Strauss, and P. R. Levitt, 1994, Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the Geostack technique: Geophysics, **59**, 1362–1376, doi: 10.1190/1.1443695.

Figueiredo, M. A. T., R. D. Nowak, and S. J. Wright, 2007, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems: IEEE Journal of Selected Topics in Signal Processing, **1**, 586–597, doi: 10.1109/JSTSP.2007.910281.

Hennenfent, G., E. van den Berg, M. P. Friedlander, and F. J. Hermann, 2008, New insights into one-norm solvers from the Pareto curve: Geophysics, **73**, no. 4, A23–A26, doi: 10.1190/1.2944169.

Herrmann, P., and G. Cambois, 2001, Statistical properties of seismically derived AVO attributes: 71st Annual International Meeting, SEG, Expanded Abstracts, 231–234.

Ikelle, L. T., and L. Amundsen, 2005, Introduction to petroleum seismology: Investigations in geophysics: SEG.

Jin, S., G. Cambois, and C. Vuillermoz, 2000, Shear-wave velocity and density estimation from PS-wave AVO analysis: Application to an OBS dataset from the North Sea: Geophysics, **65**, 1446–1454, doi: 10.1190/1.1444833.

Larson, R., and B. H. Edwards, 1999, Elementary linear algebra, 4th ed.: Houghton Mifflin Company.

Misra, S., and M. D. Sacchi, 2008, Global optimization with model-space preconditioning: Application to AVO inversion: Geophysics, **73**, no. 5, R71–R82, doi: 10.1190/1.2958008.

Oldenburg, D. W., T. Scheuer, and S. Levy, 1983, Recovery of the acoustic impedance from reflection seismograms: Geophysics, **48**, 1318–1337, doi: 10.1190/1.1441413.

Ostrander, W. J., 1984, Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence: Geophysics, **49**, 1637–1648, doi: 10.1190/1.1441571.

Pérez, D. O., and D. R. Velis, 2011, Sparse-spike AVO/AVA attributes from prestack data: 81st Annual International Meeting, SEG, Expanded Abstracts, 340–344.

Ricker, N., 1940, The form and nature of seismic waves and the structure of seismograms: Geophysics, **5**, 348–366, doi: 10.1190/1.1441816.

Robinson, E. A., and D. A. Ebrom, eds., 1996, Deconvolution 2: Geophysics Reprint Series 17: SEG.

Robinson, E. A., and S. Treitel, 2002, Geophysical signal analysis: SEG.

Sacchi, M. D., 1997, Reweighting strategies in seismic deconvolution: Geophysical Journal International, **129**, 651–656, doi: 10.1111/j.1365-246X.1997.tb04500.x.

Sacchi, M. D., D. R. Velis, and T. J. Ulrych, 1996, Wavelet via polycepstra: 66th Annual International Meeting, SEG, Expanded Abstracts, 1583–1586.

Shuey, R., 1985, A simplification of the Zoeppritz equations: Geophysics, **50**, 609–614, doi: 10.1190/1.1441936.

Smith, G. C., and M. Gidlow, 1987, Weighted stacking for rock property estimation and detection of gas: Geophysical Prospecting, **35**, 993–1014, doi: 10.1111/j.1365-2478.1987.tb00856.x.

Smith, G. C., and M. Gidlow, 2000, A comparison of the fluid factor with λ and μ in AVO analysis: 70th Annual International Meeting, SEG, Expanded Abstracts, 122–125.

Tarantola, A., 1987, Inverse problem theory, methods for data fitting and model parameter estimation: Elsevier Science Publishing Co.

Taylor, H. L., S. C. Banks, and J. F. McCoy, 1979, Deconvolution with the l_1-norm: Geophysics, **44**, 39–52, doi: 10.1190/1.1440921.

Ulrych, T. J., and M. D. Sacchi, 2005, Information-based inversion and processing with applications: Elsevier.

van den Berg, E., and M. Friedlander, 2009, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912, doi: 10.1137/080714488.

Velis, D. R., 2003, Estimating the distribution of primary reflection coefficients: Geophysics, **68**, 1417–1422, doi: 10.1190/1.1598135.

Velis, D. R., 2008, Stochastic sparse-spike deconvolution: Geophysics, **73**, no. 1, R1–R9, doi: 10.1190/1.2790584.

Walden, A., and J. Hosken, 1986, The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution: Geophysical Prospecting, **34**, 1038–1066, doi: 10.1111/j.1365-2478.1986.tb00512.x.

Whitcombe, D. N., M. Dyce, C. J. S. McKenzie, and H. Hoeber, 2004, Stabilizing the AVO gradient: 74th Annual International Meeting, SEG, Expanded Abstracts, 232–235.

Wiggins, R., 1978, Minimum entropy deconvolution: Geoexploration: International Journal of Mining and Technical Geophysics and Related Subjects, **16**, 21–35, doi: 10.1016/0016-7142(78)90005-4.

Yilmaz, O., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Investigations in geophysics: SEG.

Zoeppritz, K., 1919, Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeits-flächen: Nachrichten der Königlichen Gesellschaft der wissenschaften zu Göttingen: Mathematish-physkalishe Klasse, **K1**, 66–84.