



Simultaneous Sparse Deconvolution

Erick Talarico, Andre Gomes, Matheus Sobreira, Carlos Cunha

Petrobras – Petróleo Brasileiro S.A

Copyright 2021, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 17th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 8-11 November 2021.

Contents of this paper were reviewed by the Technical Committee of the 17th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

Abstract

We present a new method to perform simultaneous sparse spike deconvolution in a group of volumes associated with different reflection angles. The proposed algorithm enforces co-localization of the spikes on the estimated reflectivity traces and allows user control of the sparsity via hyperparameters. The method is validated in both synthetic and real datasets with positive results. The resulting angle-dependent reflectivity volumes are suited to be used as input to perform high resolution elastic inversion.

Introduction

The band-limited character of routine seismic image volumes imposes resolution constraints for hydrocarbon reservoir characterization. These constraints are particularly restrictive for depositional settings with complex stratigraphic features.

Based on the assumption that the subsurface rocks are mainly composed of layers with sharp boundaries, sparse spike deconvolution methods extend the spectral bandwidth of these volumes, thus improving the resolution of heterogeneous stratigraphic features (Hargreaves et al., 2013; Kazemi and Sacchi, 2014; Rosa, 2018). The general goal of these methods is to transform the seismic traces in an estimate of the reflection coefficient series associated with the layer boundaries.

Acoustic inversion methods can also benefit from the bandwidth extension provided by sparse spike deconvolved data, by using it as input to the inversion software (Cunha et al., 2018).

In order to extend the applicability of sparse spike deconvolution to elastic inversion, it is advisable to constraint the solutions from traces associated with the same spatial position and different reflection angles, so that the angle-dependent reflectivity's are representative of the same geological boundaries. (Xi et al., 2018) solves the problem by using a multivariate version of a modified Cauchy distribution. But, unlike the solution presented in this paper, the modified Cauchy distribution in (Xi et al., 2018) has no parameter to control the sparsity of the solution and does not develop the full posterior distribution over the reflectivities.

Our work is based on Automatic Relevance Determination (ARD) formalism explained in (Bishop, 2006). This is a fully Bayesian formalism which was created for linear regression problems, where one is interested in finding the sparsest solution (solution with a minimal number of non-zero coefficients), but it can be applied to any linear inverse problem. In the geophysics literature, there are few publications that explore ARD formalism: (Valentine and Sambridge, 2018) present it in the context of inversion regularization techniques; (Ji et al., 2020) applies ARD to AVA joint inversion to P and S reflectivities.

The novelty of the present work is that we apply the ARD algorithm to AVA inversion with a different kind of prior than the one used in (Ji et al., 2020), which enforces co-localization of the reflectivity spikes throughout the different reflectivity traces. The present paper also makes no assumption regarding the seismic error spatial structure, which is determined in the proposed optimization algorithm.

The results presented in this paper discuss the deconvolution of seismic stacks problem, although the tools developed here could also be applied to any kind of AVA seismic inversion (such as AVA to elastic reflectivities inversion for example).

Theory

In this section, the general theory for ARD will be introduced, and its application formulas for seismic deconvolution will be presented.

ARD applied to geophysics

Many geophysical problems may be cast into or approximated by a linear system, with gaussian noise. For example, given an input vector r , and an output vector s , with the linear transformation G , one can write:

$$\begin{aligned} s &= Gr + \epsilon \\ \epsilon &\sim \mathcal{N}(\epsilon|0, S) \end{aligned} \quad (1)$$

Where S is the covariance matrix of the measurement noise, and is usually assumed to be diagonal and uniform (independent and identically distributed noise). In this work, we will make no such assumption.

For seismic 1D convolutional model the matrix G is a given Toeplitz matrix, which represent the convolution of the input reflectivity trace r with the estimated seismic wavelet. The prior probability over r is defined by:

$$p(r) = \mathcal{N}(r|0, A^{-1}) \quad (2)$$

The ARD assumption is that the prior precision matrix A is diagonal, with independent entries:

$$A = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_N \end{bmatrix} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) \quad (3)$$

$$= \text{diag}(\lambda)$$

In a maximum a posteriori (MAP) approach, one is interested in finding the vector r which maximizes the posteriori probability $p(r|s, \lambda, S)$:

$$p(r|s, \lambda, S) = \frac{p(s|r, S)p(r|\lambda)}{p(s|\lambda, S)} \quad (4)$$

The above maximization problem is equivalent to minimizing the loss:

$$U(r) = (s - Gr)^T S^{-1} (s - Gr) + \sum_{j=1}^N \lambda_j r_j^2 \quad (5)$$

Which is equivalent to a deterministic inversion with regularization term. In this setting the best choice of λ from the mismatch point of view is zero (the smaller the regularization, the better is the match between measured s and simulated Gr). On the other hand, the smaller mismatch usually comes at the expense of unstable (noisier) solutions. Figure 1 illustrates this situation for a 2D toy example. The Bayesian approach brings a natural way of finding the optimal value of the regularization hyperparameters λ that balances mismatch and solution quality.

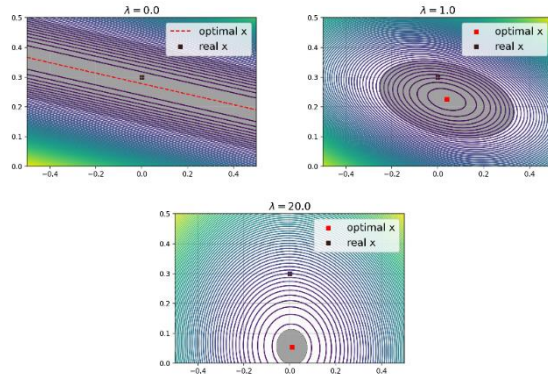


Figure 1: Bayesian inversion applied to a 2D toy example for different values of the hyperparameter $\lambda_1 = \lambda_2 = 0$. (left), 1.0 (right), and 20.0 (below). The matrix G is singular, so when $\lambda_1 = \lambda_2 = 0$ there are infinite optimal solutions.

In a fully Bayesian approach, one is interested in characterizing the posterior distribution $p(r|s, \lambda, S)$; and in this context, the optimal parameters (λ, S) are obtained by maximizing the marginal evidence $p(s|\lambda, S)$:

$$p(s|\lambda, S) = \int_r p(s|r, S)p(r|\lambda) dr \quad (6)$$

The above integral can be regarded as an inner product of the functions $f_s(r) = p(s|r, S)$ and $g_\lambda(r) = p(r|\lambda)$. For a fixed seismic error matrix S , the maximum marginal evidence is achieved by finding the vector λ which maximizes the projection of $g_\lambda(r)$ over $f_s(r)$. This is neither achieved with $\lambda \approx \vec{0}$ (equiprobable distribution over r) nor with $\lambda \rightarrow \infty$ (distribution of r peaked at the solution $r = \vec{0}$); but there is an optimal non-trivial solution.

The above integral is not analytical to be computed and its derivatives are also not analytical (and they suffer from vanishing gradients). For this reason, one can use the identity below, which holds for any test distribution $q(r)$:

$$\log p(s|\lambda, S) = \mathcal{L}(q, \lambda, S) + D_{KL}(q(r)|p(r|s, \lambda, S)) \geq \mathcal{L}(q, \lambda, S) \quad (7)$$

$$\mathcal{L}(q, \lambda, S) = \mathbb{E}_{q(r)} \left[\log \frac{p(s, r|\lambda, S)}{q(r)} \right] \quad (8)$$

$$D_{KL}(q(r)|p(r|s, \lambda, S)) = \mathbb{E}_{q(r)} \left[\log \frac{q(r)}{p(r|s, \lambda, S)} \right] \quad (9)$$

The D_{KL} is the Kullback-Leibler divergence and measures the discrepancy between the proposed distribution $q(r)$ and the true posterior $p(r|s, \lambda, S)$. It is always greater than or equal to zero; equality only holds when $q(r) \equiv p(r|s, \lambda, S)$.

The $\mathcal{L}(q, \lambda, S)$ term is known as Evidence Lower Bound and maximizing it with respect to q is equivalent to minimizing the D_{KL} term. The EM algorithm (Expectation Maximization) consists of optimizing the marginal evidence $p(s|\lambda, S)$ by alternating two optimization steps. It can be described in the Algorithm 1:

Input:	seismic trace s , number of iterations n_{iter}
Output:	optimum reflectivity precisions $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$ and measurement error covariance matrix S
1	Initialize λ_0 and S_0
2	for $i = 1$ to n_{iter} do:
3	E-step: set the test distribution $q(r) = p(r s, \lambda_0, S_0)$
4	M-step: maximize the ELBO $\mathcal{L}(q, \lambda, S)$ in relation to λ and S , fixing $q(r) = p(r s, \lambda_0, S_0)$
5	set $\lambda_0 = \lambda$ and $S_0 = S$
6	set $\lambda_0 = \lambda$ and $S_0 = S$
7	end

Algorithm 1: Expectation Maximization (EM) algorithm for maximizing the marginal evidence probability.

The posterior distribution of the E-step is developed in (Buland and Omre, 2003), and can be written as:

$$p(r|s, \lambda_0, S_0) = \mathcal{N}(r|\hat{r}, \Lambda_r^{-1}) \quad (10)$$

$$\hat{r} = \Lambda_r^{-1} G^T S_0^{-1} s$$

$$\Lambda_r = A_0 + G^T S_0^{-1} G$$

In the M-step, one can ignore the terms which do not depend on λ or S :

$$\lambda, S = \underset{\lambda, S}{\operatorname{argmax}} \mathbb{E}_{q(r)} [\log p(s, r|\lambda, S)] \quad (11)$$

$$= \underset{\lambda, S}{\operatorname{argmax}} \int_r p(r|s, \lambda_0, S_0) \log p(s, r|\lambda, S)$$

The solution to the above optimization problem is given by:

$$\lambda_i = \frac{1}{[\Lambda_r^{-1}]_{i,i} + \hat{r}_i^2} \quad (12)$$

$$S = G \Lambda_r^{-1} G^T + (G \hat{r} - s)(G \hat{r} - s)^T \quad (13)$$

The above equations 10 through 13 are recursive equations since they have cyclic dependencies. The solution is found, thus by iterative application of these equations.

In practice the seismic noise covariance matrix in Equation 13 is rescaled at each iteration to match the desired signal

to noise ratio, which must be inputted by the expert. The reason is that the seismic noise is in the same bandwidth as the signal and, therefore, much of the noise is indistinguishable from the signal and wouldn't be correctly estimated by Equation 13. As such, the signal to noise parameter indirectly controls the sparsity of the solution: for higher signal to noise ratio the algorithm will try to match each and every seismic event, and the solution will be less sparse; for smaller signal to noise ratio, more events will be ignored by the algorithm, which will be able, then to yield a sparser solution.

The EM algorithm tries to find the smallest values for the vector λ , and by doing that it induces sparsity in the solution (if consistent with the matrix G): the necessary entries of the vector r will be associated with smaller values λ_i , while the entries of r which can be set to zero, without compromising the measurement mismatch, will be associated with high λ_i values, and set to approximately zero after some iterations. Figure 2 exemplifies the deconvolution of a seismic trace in the ARD formalism: the measured seismic equals the true seismic plus a coloured noise; the ARD solution has a good match with the true reflectivity trace, is sparser than the conventional deconvolution with similar seismic mismatch. Figure 3 compares the estimated and actual seismic noise covariance functions and shows that they are similar.

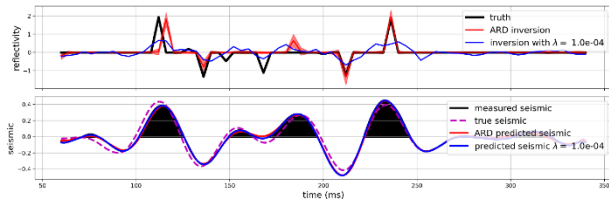


Figure 2: Application of the ARD formalism to deconvolve a seismic trace. It can be noted that the solution is sparse, and it has a good match with the measured seismic. An inversion with $\lambda_1 = \lambda_2 = \dots = \lambda_N = 10^{-4}$ resulted in a not sparse solution with the same mismatch level.

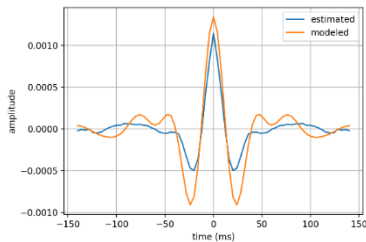


Figure 3: Covariance function of the seismic noise as a function of time lag. There is a good match between the estimated and modeled error structures.

ARD for multistack deconvolution

In a multistack deconvolution, the matrix G is a block-diagonal matrix, where each block is the Toeplitz matrix associated with each stack respective wavelet. We would like to induce sparsity, without compromising correlation between the different stack reflectivities, since their correlation is a direct consequence of the rock physics correlation between the different elastic properties.

The way to do that is to elaborate more on the structure of the prior precision matrix A . Say, we choose A to be:

$$A = A_\theta \otimes A_t \quad (14)$$

$$A_\theta = \text{diag}(\lambda_{\theta_1}, \lambda_{\theta_2}, \dots, \lambda_{\theta_F}) = \text{diag}(\lambda_\theta) \quad (15)$$

$$A_t = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) \quad (16)$$

Where \otimes is the kronecker product, F is the number of stacks, and N is the number of trace samples. The above equations mean that a sample at position i and stack j has precision equals to $\lambda_{\theta_j} \lambda_i$. This means that the matrix A_θ controls the weight of each stack in the solution and is shared among all samples. Since each reflectivity trace is important to explain its correspondent seismic stack trace, the optimized A_θ will not induce sparsity, since it would mean to zero a whole reflectivity trace. Most importantly, equations 14 through 16 imply that if a sample is not important to explain the output seismic gather, then $\lambda_i \gg 1$, and therefore, all stack reflectivities will have $\lambda_{\theta_j} \lambda_i \gg 1$, meaning that the sparsity will be shared across the different stack reflectivities.

The EM algorithm, together with the above prior precision matrix assumption is what we will refer to as the simultaneous sparse deconvolution (SSD) formalism.

Results

We will show the application of the SSD formalism to both synthetic and a real case example. The SSD formalism will be compared to the application of the ARD approach independently to each stack, which we will refer to as independent sparse deconvolution (ISD).

Synthetic case

We have created a sparse reflectivity trace by sampling from a Gaussian-Bernoulli process (sampling a random white noise and multiplying it by a binary random trace with 0.1 probability to be equal to 1). We have derived elastic reflectivity traces (P, S and density) proportional to this initial reference reflectivity and guided by linearized rock physics relations. The simulated stacks correspond to the reflection angles of 0, 15 and 30 degrees (Near, Mid and Far respectively), and their respective reflectivities were computed using Aki-Richards equations. The wavelets were modelled as butterworth filters with the same low-cut ramp (5Hz-10Hz) but decreasing highcut ramps: 50Hz-80Hz for the Near stack, 30Hz-50Hz for the Mid stack, and 20Hz-40Hz for the Far stack. A colored noise with the same power spectrum as the wavelets was added such that the signal to noise ratio equals 20.

Figure 4 shows the deconvolution results for both ISD and SSD formalisms. It can be seen that the ISD has no one-to-one correspondence between spikes across the different stacks and no alignment is guaranteed, whereas the SSD has aligned spikes and better match with the true reflectivities. Since ARD is a fully Bayesian formalism, Figure 4 also illustrates the 80% confidence interval of the reflectivities for the SSD case.

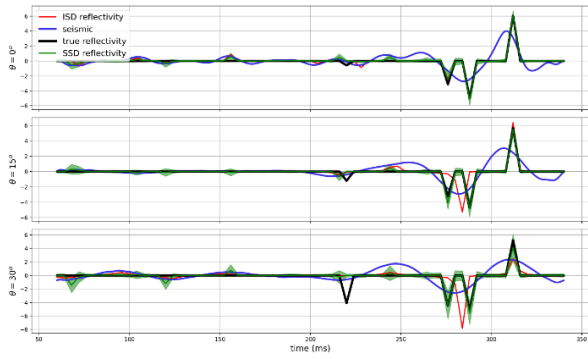


Figure 4: Comparison of ISD and SSD results. Each plot represents the deconvolution of a seismic stack (0, 15 and 30 degrees). The independent deconvolution (red traces) shows no one-to-one correspondence between the spikes across the different stacks, and no alignment is guaranteed. The simultaneous deconvolution (green traces) shows better correlation between spikes in different stacks, and better match with the true reflectivity traces. For the SSD case, the 80% confidence interval is also displayed.

Real Case

The real case example was taken from an offshore seismic. The wavelets were calibrated from nearby wells (although statistical wavelets could also be used). The signal to noise ratio was qualitatively calibrated respectively as 5, 5, 2, 1 for the Near, Mid, Far and Ultra Far stacks (alternatively, those values can be estimated from seismic to well tie), meaning that the Far and Ultra Far stacks are much more deteriorated by noise than the Near and Mid.

A first comparison between the ISD and SSD approaches is illustrated in Table 1, which presents the correlations between different reflectivity stacks.

	Near	Mid	Far	Ultra Far
Near				
Mid	0.339			
Far	0.111	0.188		
Ultra Far	0.015	0.038	0.118	

Table 1: Correlation coefficients between the different reflectivity stacks for the SSD (green) and for the ISD (orange) cases. It can be noticed that the SSD correlation coefficients tend to have higher values.

Figure 5 compares the deconvolution results for a random trace within the seismic volume. It confirms visually that the SSD approach yields more aligned reflectivity traces.

Visual analysis of the 80% confidence intervals in Figure 5 shows a tendency of increasing uncertainty from the near stack (7 degrees reflection angle) to the ultra-far stack (36 degrees reflection angle).

Figure 5 also shows some misalignment issues in the real seismic data, for example, around 5860 ms. These misalignments impact the stability of the SSD solution: the amplitude of the reflectivity spikes is bigger for the far and ultra-far angle stacks, and they tend to present a dipole-like behavior aspect at the misaligned seismic events.

Figure 6 illustrates SSD on a real seismic line. The misalignment issues between Near and Ultra-far stacks is clear, but still there is great structural conformity between Near and Ultra-far reflectivities as can be seen from the sections.



Figure 5: Result of the deconvolution on a single trace of the real seismic data.

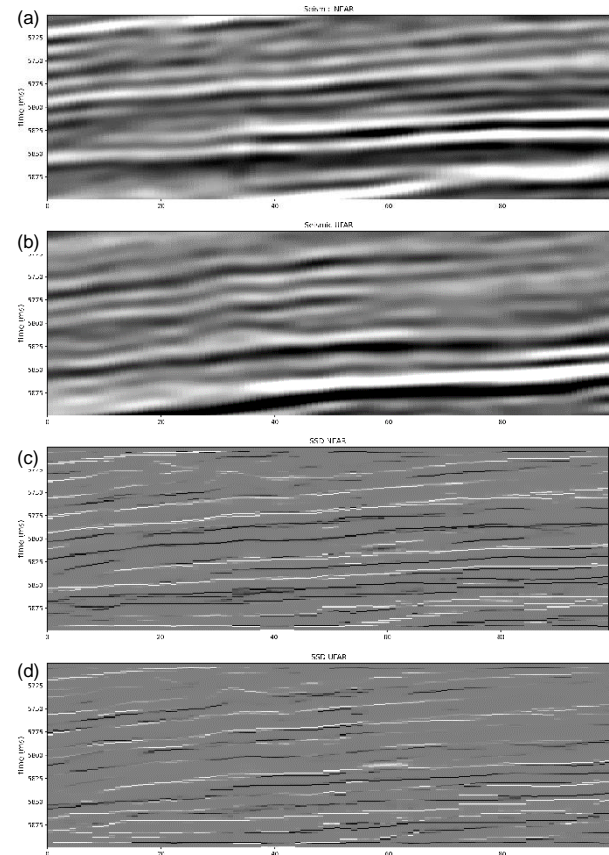


Figure 6: SSD inversion results. (a) Near stack seismic. (b) Ultra-far stack seismic. (c) Near reflectivity. (d) Ultra-far reflectivity.

Conclusions

The presented SSD (Simultaneous Sparse Deconvolution) technique developed here is based on the ARD formalism, together with a specific structure in the prior precision matrix, which couples the sparsity of different reflectivity traces. It has been shown, both in synthetic and real case examples, that the SSD technique enforces co-localization between spikes, resulting in more accurate results and better correlation between the resulting reflectivity traces.

Acknowledgments

The authors would like to acknowledge Petrobras for the support and permission to publish the present work.

References

Cunha, C. A., A. Damasceno, A. Pimentel, L. Teixeira, N. Cruz, T. Oliveira, High Resolution Impedance Inversion: Brazilian Journal of Geophysics VOL 37, No 4, 461-469, DOI: <http://dx.doi.org/10.22564/rbgf.v37i4.2022>

Hargreaves, N., S. Treitel, and M. Smith, 2013, Frequency extension, resolution, and sparse inversion: 83th SEG Annual Meeting Expanded Abstracts, DOI: <http://dx.doi.org/10.1190/segam2013-0782.1>

Kazemi, N., and D. Sacchi, 2014, Sparse multichannel blind deconvolution: Geophysics, 79, no. 5, V143–V152, DOI: <http://doi.org/10.1190/geo2013-0465.1>

Rosa, A. L. R., 2018, The seismic signal and its meaning: Published by Society of Exploration Geophysics, DOI: <https://doi.org/10.1190/1.9781560803348>

Xi, Y., X. Yin and Z. Zong, 2018, AVO multi-trace group sparse inversion: 88th SEG Annual Meeting Expanded Abstracts, DOI: [10.1190/segam2018-2995959.1](https://doi.org/10.1190/segam2018-2995959.1)

Bishop, C., 2006, Pattern Recognition and Machine Learning: Published by Springer, ISBN-10: 0387310738

Sroubek, F., Smidl, V., Kotera, J., 2014, Understanding image priors in blind deconvolution: 2014 IEEE International Conference on Image Processing (ICIP), DOI: [10.1109/ICIP.2014.7025911](https://doi.org/10.1109/ICIP.2014.7025911)

Valentine, A., Sambridge, M., 2018, Optimal regularization for a class of linear inverse problem: Geophysical Journal International, DOI: <https://doi.org/10.1093/gji/ggy303>

Ji, Y., Hu, H., Lin, Z., Zhang, K., Zhong, H. 2020, The preconditioned ARD-based AVA inversion method for P-impedance and S-impedance: SEG International Exposition and 90th Annual Meeting, DOI: <https://doi.org/10.1190/segam2020-3415458.1>

Buland, A., Omre, H., 2003, Bayesian linearized AVO inversion: Geophysics, 68, no. 1, 185-198, DOI: [10.1190/1.1543206](https://doi.org/10.1190/1.1543206)