APPLICATION OF THE L-CURVE IN SEISMIC TRAVELTIME TOMOGRAPHY: METHODOLOGIES FOR THE EXTRACTION OF THE OPTIMAL REGULARIZATION PARAMETER

Felipe Antonio TERRA[1]
Amin BASSREI[2]
Eduardo Telmo Fonseca SANTOS[3]

[1] Bacharel em Geofísica, Mestre em Geofísica. Geofísico, Petrobras S/A. felipeterra@petrobras.com.br
[2] Engenheiro Eletricista, Dr. em Geofísica Aplicada. Professor Associado, Depto. de Geofísica, Instituto de Geociências, Centro de Pesquisa em Geofísica e Geologia, Universidade Federal da Bahia – UFBA; Instituto Nacional de Ciência e Tecnologia de Geofísica de Petróleo. bassrei@ufba.br
[3] Bacharel em Ciência da Computação, Mestre em Engenharia Elétrica, Dr. em Geofísica Aplicada. Professor Adjunto, Depto de Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Bahia – IFBA. eduardo.telmo@terra.com.br

ABSTRACT. Inverse problems in Applied Geophysics are usually ill-posed. One way to reduce such deficiency is through derivative matrices, which are a particular case of a more general procedure called regularization. The regularization by derivative matrices has an input parameter called regularization parameter, usually denoted by λ, which choice is already a problem. It was suggested in the 1970's a heuristic approach later called L-curve, with the purpose to provide the optimum regularization parameter. The L-curve, called this way due to the fact of having the approximate shape of the letter L, is a parametric curve, where each point is associated to a parameter λ. In the horizontal axis it is represented the inversion error, that is, the error between the observed data and the calculated one, and, in the vertical axis, it is represented the amount of regularization, that is, the product between the regularization matrix and the estimated model. The ideal point is the L-curve knee, where there is a balance between the quantities represented in the Cartesian axes. The L-curve has been applied to a variety of inverse problems, including Geophysics. However, the visualization of the knee and the extraction of the optimal parameter is not always an easy task, particularly when the L-curve does not have a clear L shape. In the present work, three methodologies are employed for the search and obtainment of the optimal regularization parameter from the L curve. The first criterion is the utilization of Hansen's toolbox, which extracts λ automatically. The second criterion consists in building and visually extracting the optimal parameter. Finally, by third criterion one understands the construction of the first derivative of the L-curve, and the posterior automatic extraction of the maximum point, which is associated to the knee in the original L-curve. The regularization theory through derivative matrices, the utilization of the L-curve, and the three above criteria were applied and validated in the classical inverse problem of traveltime tomography. After many simulations with synthetic data, noise-free as well as data corrupted with noise, with the regularization orders 0, 1, and 2, it was verified that the three criteria are valid, and in general provide satisfactory results. The third criterion presented the best performance, especially in cases where the L-curve has an irregular shape.
Key-words: inverse problems, reservoir geophysics, traveltime tomography, regularization, L-curve.

RESUMO. Os problemas inversos da Geofísica Aplicada são geralmente são mal postos. Uma maneira de reduzir essa deficiência é através de matrizes de derivadas, que são um caso particular de um procedimento mais geral denominado regularização. A regularização por matrizes de derivadas tem um parâmetro de entrada chamado parâmetro de regularização, geralmente denotado por λ, cuja escolha já é um problema. Foi sugerida na década de 1970 uma abordagem heurística mais tarde chamada de curva L, com a finalidade de fornecer o parâmetro ótimo de regularização. A curva L, denominado desta forma devido ao fato de ter a forma aproximada da letra L, é uma curva paramétrica, onde cada ponto está associado a um parâmetro λ. O eixo horizontal representa o erro de inversão, isto é, o erro entre os valores observados e os valores calculados do vetor dado. O eixo vertical representa a quantidade de regularização, isto é, o produto entre a matriz de regularização e o vetor estimado de parâmetro de modelo. O ponto de ideal é o joelho da curva L, onde existe um equilíbrio entre as quantidades representadas nos eixos cartesianos. A curva L foi aplicada a uma variedade de problemas inversos, inclusive na Geofísica. No entanto, a visualização do joelho para a extração do parâmetro ideal nem sempre é uma tarefa fácil, em especial quando a curva L não tem claramente o formato da letra L. No presente trabalho, três metodologias são empregadas para a busca e obtenção do parâmetro de regularização ideal a partir da curva L. O primeiro critério é a utilização do pacote de programas de Hansen, que extrai o λ automaticamente. O segundo critério consiste em construir a curva L, e visualmente extrair o parâmetro ótimo. Finalmente, o

terceiro critério compreende a construção da curva da primeira derivada da curva L, e a extração automática posterior do ponto máximo, o qual está associado ao joelho da curva L original. A teoria de regularização através de matrizes de derivadas e a utilização da curva L, juntamente com os três critérios acima referidos foram aplicados e validados no problema inverso clássico da tomografia de tempos de trânsito. Depois de muitas simulações com dados sintéticos, sejam dados livres de ruído como também dados corrompidos com ruído, com as ordens de regularização 0, 1 e 2, verificou-se que os três critérios são válidos e os resultados satisfatórios. O terceiro critério apresentou o melhor desempenho, especialmente nos casos em que a curva L tem uma forma irregular.
Palavras-chaves: problemas inversos, geofísica de reservatórios, tomografia de tempos de trânsito, regularização, curva L.

INTRODUCTION

Exploration seismology or simply seismics is the Geophysics branch that is more often applied to image the subsurface for the oil industry. It uses several techniques based on the theory of propagation of elastic and acoustic waves and its laws of reflection and refraction. Tomography is a technique originally applied in medical imaging and since 1970's it has been used in Geophysics for electromagnetic and seismic imaging. In the particular case of seismic tomography, it is an inversion procedure that provides images with higher resolution than those provided by conventional reflection seismology. In this work, we use traveltime tomography, where the input data (data parameter vector **d**) are the traveltimes between sources and receivers, and the output data (model parameter vector **m**) is the slowness distribution of the 2-D discretized medium.

As an ill-posed inverse problem we have here the classical issues of existence, uniqueness and stability. These difficulties are mostly due to the fact that the kernel matrix is ill-conditioned, and the solutions are sensitive to the perturbations in the data. Additional difficulties are the presence of noise, the lack of information and the fact that our model is discrete. In order to attenuate this deficiency, it is necessary some regularization technique. In this work, we used the regularization by derivative matrices, usually known in the literature as Tikhonov regularization. This technique needs the selection of a regularization parameter or factor, usually expressed by λ which choice is a relevant problem by itself, if feasible solutions are desired. During many years the search of the optimum λ was done by trial and error. In this work, we use the L-curve for the selection of the optimum λ, which was reintroduced in the literature by Hansen (1992).

The L curve, so called this way because very often it takes the form of that letter of the alphabet, is a parametric curve, where each point is associated with a parameter λ. The horizontal axis displays the inversion error, that is, the difference between the observed data and calculated data, and in the vertical axis it is represented the amount of regularization, which is the product between the regularization matrix and estimated solution. The ideal point would then be the curve knee, where there is a balance between the two quantities represented in Cartesian axes. The L curve has found application in a variety of inverse problems, including Geophysics. However, the identification of the knee and the subsequent extraction of the optimum parameters are not always trivial tasks, especially when the L curve does not have the clear shape of the letter L.

Ray and Sanches (1994) introduced an early work in Geophysics using the L curve for the choice of λ in order to obtain tidal estimates based on altimetry data from Geosat satellite. Sá (1996) also used the L curve in traveltime tomography, and Bassrei and Santos (2007) used the L curve in geophysical diffraction tomography.

The main objective of this work is to present three different methods for the extraction of the optimal parameter λ from the curve L: (i) the Hansen toolbox, (ii) the method of visual inspection of the L curve, and (iii) the method of the discrete first derivative of the L curve. We performed several simulations with synthetic data in traveltime tomography for different regularization orders.

After many simulations with synthetic data, both with noise-free data and also with data contaminated by noise, in regularization orders 0, 1, and 2, it was verified that these three criteria are valid and usually provide satisfactory results. However the third criterion showed a better performance, especially in cases where the L curve has an irregular shape.

GENERALIZED INVERSE AND SINGULAR VALUE DECOMPOSITION

Consider a modeling process where the input of some system is described by certain parameters contained in **m** and the output is described as $\mathbf{d} = A\mathbf{m}$, which is a linear transformation on **m**. If the vector **d** describes the observed output of the system, the problem is reduced to "choose" the parameters **m** in order to minimize in some sense, the difference between the observed **d** and the prescribed output of the system A **m**. If we measure

this difference through the norm ||•||, our task is to find the value of **m** which minimizes
$|| A \mathbf{m} - \mathbf{d}||_2$,
where the *M* X *N* matrix *A* and the data vector **d** with *M* elements are provided to the problem. This is called a least squares problem, which can be formally stated as follows (1). Considering the basic relationship

$$\mathbf{d} = A\mathbf{m}, \quad (1)$$

we wish to minimize the error using the following objective function based on the work of Levenberg (1944) and Marquardt (1963):

$$\Phi(\mathbf{m}) = \mathbf{e}^T\mathbf{e} + \lambda L_2, \quad (2)$$

where the error is given by $\mathbf{e} = \mathbf{d} - A\mathbf{m}$, λ is a scalar called the damping parameter and $L_2 = \mathbf{m}^T\mathbf{m}$. The estimated solution, also called damped least squares solution, is

$$\mathbf{m}^{est} = (A^T A + \lambda I)^{-1} A^T \mathbf{d}. \quad (3)$$

Generalized inverse (GI) is frequently used in the inversion of geophysical data and its respective solution has the minimum norm. In this case, the objective function to be minimized is

$$\Phi(\mathbf{m}) = \mathbf{m}^T\mathbf{m} + \mathbf{t}^T(\mathbf{d} - A\mathbf{m}), \quad (4)$$

where **t** is the vector of Lagrange multipliers. The minimization yields

$$\mathbf{m}^{est} = A^T(AA^T)^{-1}\mathbf{d}. \quad (5)$$

The concept of GI was developed by Moore and also independently by Penrose (1955). Consider a *M* X *N* matrix *A*. If the following conditions are satisfied: (i) $AA^+A = A$, (ii) $A^+AA^+ = A^+$, (iii) $(AA^+)^T = AA^+$, and (iv) $(A^+A)^T = A^+A$, then the *N* X *M* matrix A^+ is unique. The GI is usually calculated using SVD or singular value decomposition (Lanczos, 1961). A rectangular *M* X *N* matrix *A* with rank *k* can be decomposed as $A = U\Sigma V^T$, where *U* is the *M* X *M* matrix which contains the orthonormalized eigenvectors of AA^T, *V* is the *N* X *N* matrix which contains the orthonormalized eigenvectors of A^TA and Σ is the *M* X *N* diagonal matrix which contains the singular values of *A*, written in the decreasing order, that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$. The GI A^+ is a *N* X *M* matrix given by $A^+ = V\Sigma^+U^T$, where Σ^+ is the *N* X *M* diagonal matrix which contains the reciprocals of the non-zero singular values of *A*, so that

$$\Sigma^+ = \begin{pmatrix} E & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

and *E* is the diagonal square matrix of order *k* expressed by

$$E = \begin{pmatrix} \sigma_1^{-1} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^{-1} \end{pmatrix}.$$

In the case of square matrices with full rank, the classical inverse and the pseudo-inverse provide the same result, in such a way that the pseudo-inverse is a generalization of the classical inverse. In this work, we will use the Moore-Penrose inverse A^+, which differs from the one used by Menke (1989) who defines $A^+ = (A^TA)^{-1}A^T$ for the least-squares case.

Thus, the equations

$$\mathbf{m}^{est} = A^{-1}\mathbf{d},$$

and

$$\mathbf{m}^{est} = (A^TA)^{-1}A^T\mathbf{d},$$

can be reformulated with the direct substitution of the classical inverse by the pseudo-inverse as

$$\mathbf{m}^{est} = A^+\mathbf{d},$$

and

$$\mathbf{m}^{est} = (A^TA)^+A^T\mathbf{d}.$$

Substituting in the above equation $A = U\Sigma V^T$ and $A^T = V\Sigma^TU^T$, we obtain

$$\mathbf{m}^{est} = (V\Sigma^TU^TU\Sigma V^T)^+V\Sigma^TU^T\mathbf{d},$$

or

$$\mathbf{m}^{est} = (V\Sigma^T\Sigma V^T)^+V\Sigma^TU^T\mathbf{d},$$

since *U* is orthogonal, that is, $U^TU=I$. The last expression can be written as

$$\mathbf{m}^{est} = (V^T)^+\Sigma^+(\Sigma^T)^+V^+V\Sigma^TU^T\mathbf{d},$$

or

$$\mathbf{m}^{est} = (V^T)^+\Sigma^+U^T\mathbf{d},$$

since $V^+V=I$ and $(\Sigma^T)^+\Sigma^T = I$. Finally

$$\mathbf{m}^{est} = V\Sigma^+U^T\mathbf{d}, \quad (6)$$

since $V^TV=I$ and $V^+V=I$ so that $(V^T)^+=V$.

REGULARIZATION

Least-squares solutions may not provide good solutions or they do not even exist sometimes. In order to solve this problem, we use a tool for regularization or smoothing: the ill-conditioning of the matrix *A* is regularized and the unstable least-squares estimate $\mathbf{m}^{est}$ is consequently smoothed to greatly reduce the possibility of wild noise-induced

fluctuation in **d**, hopefully without distorting the resulting smoothed image too far from the true **m** (Titterington, 1985). The concept of regularization was introduced by Tikhonov in 1963 in order to improve the quality of the inversion. Many researchers studied this theory and we use the Twomey (1963) approach. For further information, please check Bassrei and Rodi (1993) for a little bit more about names and history in regularization theory. Consider the following objective function:

$$\Phi(\mathbf{m}) = \mathbf{e}^T\mathbf{e} + \lambda(D_l\mathbf{m})^T(D_l\mathbf{m}), \quad (7)$$

where λ is the regularization parameter and D_l is the l-order derivative matrix. If $\partial\Phi(\mathbf{m})/\partial\mathbf{m} = 0$, then the estimated model is given by

$$\mathbf{m}^{est} = (A^TA + \lambda D_l^T D_l)^{-1}A^T\mathbf{d}, \quad (8)$$

Notice that if λ=0 we obtain the standard least squares. The least squares is said to be damped if $D_0^T D_0 = I$. If D is the first derivative matrix then the regularization is called to be first order and so on. The matrices D_1 and D_2 are expressed as follows:

$$D_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & \cdots & \ddots & \vdots & & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix},$$

and

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & \cdots & \ddots & \vdots & & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

In the last 25 years or so some researchers studied the problem of finding the optimum λ in geophysical applications. One of first works was done by Ray and Sanchez (1994), who used regularization and L-curve to raw tidal estimates based on Geosat altimeter data. The estimation is based on fitting using specific functions called Proudman functions as spatial basis, which is formulated as a linear system. The coefficients of fitting are obtained solving the corresponding least-squares problem using zero-order regularization. Then, L-curve is applied to obtain the optimal regularization parameter.

Yao and Roberts (1999) presented an algorithm for the practical choice of the regularization parameter in linear seismic tomographic inversion. Two criteria for the choice of the regularization parameter are investigated. The first approach assumes that norm of the errors in observed data is known accurately and searches the regularization parameter associated with this error using Newton's method. The second approach is the application of generalized cross-validation (GCV) which chooses the regularization parameter associated with the best average prediction for all possible omissions of one datum, corresponding to the minimizer of GCV function.

More recently, Farquharson and Oldenburg (2004) compared two automatic ways of estimating the best regularization parameter to non-linear inverse problems: GCV and L-curve. These criteria initially proposed for linear problems are applied to the each iteration of the linearized inverse problem, in a typical iterative process to obtain the linearized solution to the corresponding non-linear problem. Thus, the best λ is estimated for each linearized iteration. To ensure that the regularization parameter decreases along iterations, an attenuation factor is multiplied by the regularization parameter from the last iteration to limit the next maximum allowable parameter.

In the L-curve, the x axis represents the error between the observed data and the calculated one, and the y axis represents the amount of regularization of the solution. L-curve was reintroduced in the literature of inverse problems by Hansen (1992a, 1998) and he also developed a software toolbox (1992b). Hansen's book (1998) is a very good source of information for a more rigorous treatment of L-curve and also mentions the pioneering contributions in this field.

The L-curve knee represents a trade-off between smoother solutions with higher errors and rougher solutions with smaller errors. Thus, the knee detection at the L-curve is a heuristic criterion to select the most appropriate solution. Solutions near to the curve knee are also acceptable and possibly more physically meaningful. We applied the L-curve implementing an automatic method to initially select the best regularization parameter, but solutions with regularization parameter near to the selected one were also considered. Thus, one can achieve a solution that simultaneously satisfies the criteria of error minimization, smoothness and also presenting physical meaningful characteristics. The detection of the L-curve corner was performed using Hansen's toolbox. Considering this curve approximately L-shaped, one can find its knee searching the maximum curvature point. However, secondary inflexions may occur, which may cause wrong detection of the best regularization parameter. Thus, the automatic method of knee detection adopted in this toolbox may lead sometimes to inadequate regularization parameters. Due to this problem, sometimes we needed to select the best regularization parameter by visual inspection of L-curve and the non-automatic detection of its knee.

We adopted a different criterion based on a curve representing the cosine of angles between adjacent segments of L-curve discrete

representation, which we named Theta curve. Wherever the curve is locally straight, the angle tends to zero, leading the cosine of this angle to one. Near the L-curve knee, the angle tends to be greater than at its neighbors, leading the cosine to values below one. Thus, smaller values of cosine are associated with inflexions of the curve, which lead us to inspect the minima of the Theta curve to find the knee of L-curve and consequently the best regularization parameter. The method developed to select the best regularization parameter is based on the detection of the first local minimum of the Theta curve. This minimum is automatically detected where the first derivative is close to zero and the second derivative is positive, adopting thresholds due to the discretization and arithmetic computer precision. Thus, the first occurrence of minimum at Theta curve is associated with the knee of L-curve, giving us the best regularization parameter. Further inflexions of the L-curve were discarded because only the first local minimum of Theta curve is associated with L-curve knee. This avoids the wrong regularization parameter detection described earlier, when one adopts the criterion of maximum curvature of the L-curve.

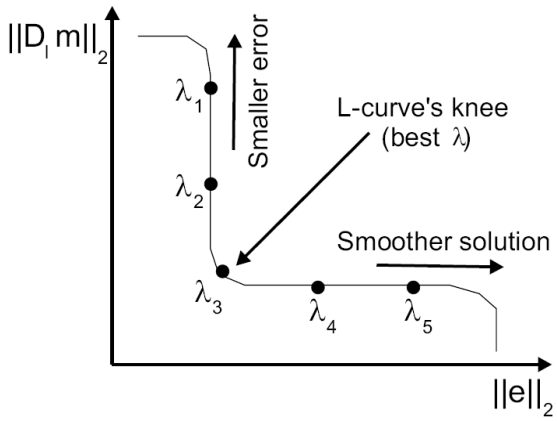
THE L-CURVE

The name L-curve comes from the fact that the curve has usually the form of the letter "L", with an inflexion point that separates the horizontal and vertical parts of the curve. According to Hansen (1998), the approximately vertical part of the curve corresponds to solutions where $||D_n\mathbf{m}||$ is very sensitive to regularization parameter changes, since the solutions are dominated by perturbation error. On the other hand, the approximately horizontal part corresponds to the solutions dominated by regularization error, since the residual norm $A\mathbf{m}-\mathbf{d}$ is more sensitive to the value of λ. The regularization parameter λ is an important scalar number that controls the regularization solution, which comes from the optimization of the objective function, that is, equation (7).

Choosing a high value for λ clearly favors a low value of $||D_n\mathbf{m}||$ and a high value of the residual norm $A\mathbf{m}-\mathbf{d}$, which is equivalent to the high amount of regularization, since the derivative of the model parameters will enforce a greater weight in the objective function, over-regularizing the solution. On the other hand, a low value for λ will provide solutions that tend to minimize the error in the data but with a low amount of regularization (under-regularization).

A good λ selection will provide the equilibrium between the error in the data and the smoothing in the model parameters. In the last 25 years a number of strategies have been suggested for the choice of λ. Hansen (1992) proposed the L-curve as a graphical tool that express the compromise of these two amounts, where the chosen model corresponds to the equilibrium situation represented by the region, known in the literature as the L-curve knee as can be seen in Figure 1. This tool was first suggested in the literature by Miller (1970) and also by Lawson and Hanson (1976). The plotting of the L-curve consists in solving a regularized linear system for each λ, computing its solution. For each estimated solution, one computes the L-curve coordinates: the discrete n-order derivative $D_n\mathbf{m}$ in the x-axis and the error $\|\mathbf{e}\| = \|\mathbf{d} - A\mathbf{m}\|$ in the y-axis. The computation of the L-curve implies an additional computation cost in the inversion procedure, but it is compensated by avoiding the time-consuming trial and error search for λ.

Figure 1 - Illustrative L-curve where $\lambda_1 < \lambda_2 < \ldots < \lambda_5$.



This obtained solution with regularization must have physical meaning, which may mean a smoother solution even if it has a slightly larger error, since the error is intrinsic to the solution and usually appears during the data acquisition process. In other words, methods that seek only the error minimization may yield inconsistent results without physical meaning. Therefore, the philosophy of the L-curve is to search a parameter that generates a balance solution between the regularization amount and the data error. The knee of the L-curve represents this balance. In this work, three methods for the extraction of the optimal parameter are presented and compared: Hansen toolbox, visual inspection of the L-curve and the first derivative of the L-curve.

Hansen's Toolbox

The toolbox called Regularization Tools, developed by P. C. Hansen in 1990-1992, has routines that are easy to use, numerically robust, based on MATLAB coding. The purpose of the toolbox is to solve ill-posed inverse problems through the Tikhonov regularization technique. The main M X N matrix A and the P X P derivative

matrix *D* are decomposed by SGVD or generalized SVD:

$$A = U\Sigma X^{-1}, \quad (9)$$

and

$$D = VSX^{-1}, \quad (10)$$

where the *M* X *N* matrix *U* and the *P* X *P* matrix *V* are orthonormal. The *P* X *P* matrix Σ contains the square root of the eigenvalues of matrix A^TA. These are called singular values σ and are disposed in the principal diagonal of matrix Σ is the decreasing order, that is, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_P$. The *P* X *P* matrix *S* contains the square root of the eigenvalues of matrix D^TD, that is, the singular values μ which are also disposed in decreasing order, that is, $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_P$. The *N* X *N* matrix *X* is non-singular. The elements of matrices Σ and *S*, which are singular values, are normalized:

$$\sigma_i^2 + \mu_i^2 = 1,$$

and the generalized singular values γ_i of the pair of matrices *A* and *D* are defined by the relationship

$$\gamma_i = \frac{\sigma_i}{\mu_i}.$$

Substituting equations (9) and (10) in equation (8) we have

$$\mathbf{m}^{est} = [(X^{-1})^T \Sigma^T U^T U \Sigma X^{-1} + \lambda (X^{-1})^T S^T V^T V S X^{-1}]^{-1} (X^{-1})^T \Sigma^T U^T \mathbf{d}.$$

Since *U* and *V* are orthonormal matrices, so that $U^TU = I$ and $V^TV = I$, the above equation can be expressed as

$$\mathbf{m}^{est} = [(X^{-1})^T \Sigma^T \Sigma X^{-1} + + \lambda (X^{-1})^T S^T S X^{-1}]^{-1} (X^{-1})^T \Sigma^T U^T \mathbf{d}$$

or

$$\mathbf{m}^{est} = [(X^{-1})^T (\Sigma^T \Sigma + \lambda S^T S) X^{-1}]^{-1} (X^{-1})^T \Sigma^T U^T \mathbf{d},$$

or

$$\mathbf{m}^{est} = X(\Sigma^T \Sigma + \lambda S^T S)^{-1} \Sigma^T U^T \mathbf{d}.$$

Using the notation $\Sigma^2 = \Sigma^T \Sigma$, $S^2 = S^T S$, and being the fact that Σ is a square matrix so that $\Sigma^T = \Sigma$:

$$\mathbf{m}^{est} = X(\Sigma^2 + \lambda S^2)^{-1} \Sigma U^T \mathbf{d}, \quad (11)$$

which can be written in vector notation as:

$$\mathbf{m}^{est} = \sum_{i=1}^{N} f_i \frac{\mathbf{u}_i^T \mathbf{d}}{\sigma_i} \mathbf{x}_i, \quad (12)$$

where the filter coefficients f_i are given by:

$$f_i = \frac{\gamma_i^2}{\gamma_i^2 + \lambda} \approx \begin{cases} 1, & \gamma_i >> \lambda, \\ \gamma_i^2 / \lambda, & \gamma_i << \lambda. \end{cases}$$

It is necessary to compute the GSVD only once. For each model estimation, which is associated to a given point of the L-curve, it is used a different set of filtered singular values. The parameter λ is related to filter intensity and when λ tends to zero it almost does not influence the division, but when it is high , it prevents small values of σ, acting thus as a damper of filter.

Since the L-curve consists of a series of discrete points, Hansen (1998) proposed to approximate it by a 2-D spline curve, and then calculate the point on the spline curve with maximum curvature. This point is defined as the knee of the L-curve. Thus, in addition to providing automatically the optimal regularization parameter from the L-curve, Hansen's package also provides the numerical treatment of ill-conditioned problems.

Visual Inspection of the L-curve

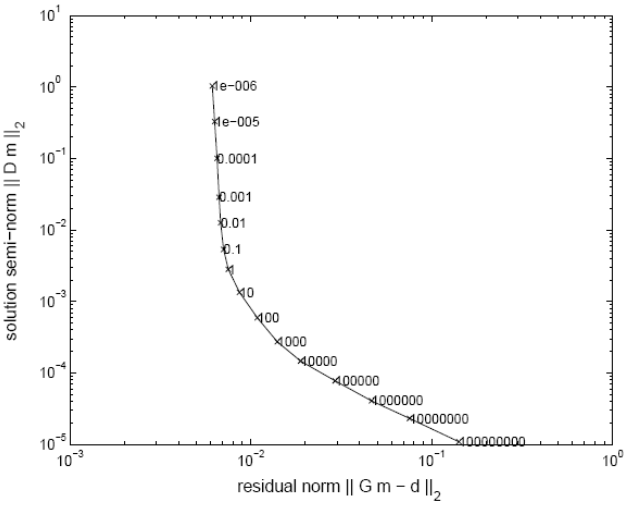
The visual inspection of the L-curve provides a non-automatic selection of the optimal regularization parameter. Even being a manual method, the visual extraction ensures a more reliable selection, since significant variations may occur in the shape of the L-curve, which can lead the automatic methods to become very susceptible to error.

This method assumes the principle of visual identification of the most obvious inflection (even of a secondary inflection may occur), the optimal regularization parameter. This inflection point corresponds to the so-called knee of the L-curve. The estimation of the inflection point of the L-curve can be made using different criteria found in the literature such as that developed by Hansen (1998), described earlier, which seeks the point of maximum curvature from the approximation of the L- curve by a 2-D spline curve. However, secondary inflection may occur in the L-curve, which will cause the incorrect detection of the optimal regularization parameter. The visual inspection may also lead to the incorrect choice, since the L-curve can present ambiguity in the choice of its inflection, as can be seen in Figure 2.

Sá (1996) proposed the L-module, which uses the nearest point of the L-curve to the origin of the curve as an estimate of the L-curve knee. But this criterion only works properly when the L-curve clearly has the shape of letter L and it may fail when the inflection knee differs significantly from a right angle, because other points outside of the knee may be closer to the origin. Aiming to reduce risks in the detection of the regularization parameter due to variations from the idealized shape of the L-curve, Santos and Bassrei (2007) proposed a different criterion, which they named as

Θ-curve, based on a curve that represents the cosine of the angles between adjacent segments of the discrete representation of L-curve. Thus, near the knee of the L-curve, the angle between adjacent segments tends to be higher than in the surrounding area, leading to cosine values less than one, with the consequent search for the minimum of the Θ-curve to find the L-curve knee.

Figure 2 - Example of a L-curve with an ambiguous inflexion point



L-curve First Derivative

The location of the inflexion point in the L-curve provides an appropriate value for the regularization parameter λ. However, the computation of the L-curve and its curvature can be computational costly for large linear systems. This is because the determination of any point on the L-curve requires both the regularized solution and the norm of the residual RMS data error. Therefore, the approximation of a L-curve through a polynomial interpolation curve can generate points of maximum curvature that are not available. This may require additional techniques. For these reasons, this paper proposes the use of the discrete first derivative of the L-curve for the automatic selection of optimal regularization factor, defined approximately as:

$$\frac{dy}{dx} \approx \frac{\left\|\mathbf{D}_n\mathbf{m}(\mathbf{e}+\Delta\mathbf{e})\right\| - \left\|\mathbf{D}_n\mathbf{m}(\mathbf{e})\right\|}{\left\|\Delta\mathbf{e}\right\|}. \qquad (13)$$

This very simple method is based on detection of the local maximum of the first derivative curve. This extreme value can be automatically detected where the derivative reaches its maximum and typically corresponds to the knee of the L-curve. In our hypothetical example, the curve in Figure 4 region is obtained from the numerical derivative of the L-curve shown in Figure 3.

Figure 3 - Example of a L-curve with a well defined inflexion point

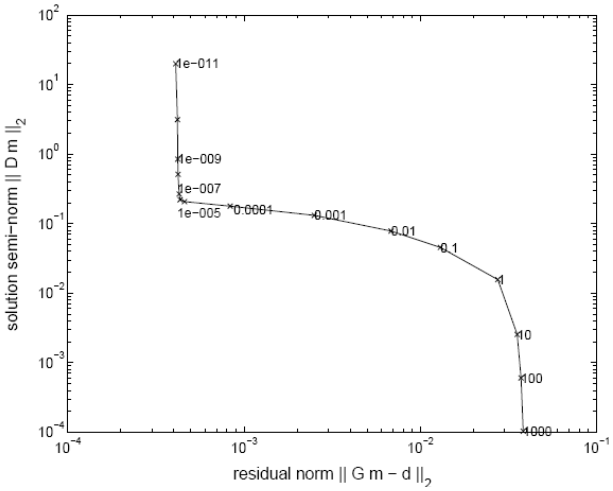
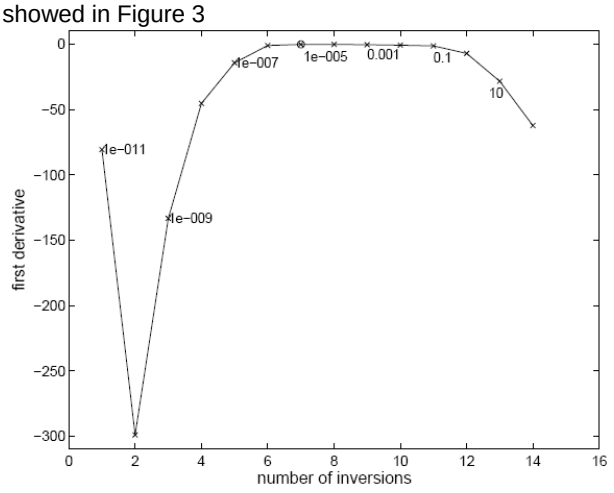


Figure 4 - Example of the first derivative of the L-curve showed in Figure 3



SEISMIC TRAVELTIME TOMOGRAPHY

Tomography is a technique of image reconstruction by the cumulative sum of property values in certain raypaths. In other words, physical parameters are estimated from projections measured at the boundary of interest region. Traveltime tomographic inversion is a special type of inverse problem that estimates object properties from line integrals.

Seismic traveltime tomography is applied to the exploration of hydrocarbons, particularly in a more detailed scale in reservoir geophysics, in order to estimate velocity models of the subsurface. The traveltimes between sources and receivers are the input data for the inversion, represented by the data vector **d**. The **G** matrix used in traveltime tomography describes the ray geometry. The parameters to be estimated, represented by the vector **m**, correspond to slowness (reciprocal of velocities), in the isotropic case, or density

normalized elastic parameters, in the anisotropic case. In this problem the vector **d** corresponds to vector of traveltimes **t**, and the vector **m** corresponds to the vector of slownesses **s**.

The traveltime t between a source and a receiver is computed through the line integral of slowness, along a raypath:

$$t_j = \int_{r_j(s)} s(x,z)dl, \quad (14)$$

where $r_j(s)$ is a raypath where the integration is performed, which is a function of slowness, *dl* is a ray element, *s(x,z)* is the slowness of the medium at coordinates *(x,z)*, where *x* is the horizontal coordinate and *z* is the vertical coordinate. The slowness model can be parameterized using a regular grid,

$$s(x,z) = \sum_{i=1}^{N} s_i B_i(x,z), \quad (15)$$

where the basis function $B_i(x,z)$ is the box function, which is 1 on the *i*-th block and zero otherwise. This model representation leads to the following relation

$$t_j \equiv g_j(\mathbf{s}), \quad (16)$$

where:

$$g(\mathbf{s}) = \sum_{i=1}^{N} s_i \int_{r_j(s)} B_i(x,z)dl. \quad (17)$$

Consider the expansion of the *j*-th traveltime observation in truncated Taylor series around $\mathbf{s}^0$:

$$t_j = t_j^0 + \left.\frac{\partial g_j}{\partial s_i}\right|_{\mathbf{s}=\mathbf{s}^0} (\mathbf{s}-\mathbf{s}^0), \quad (18)$$

or

$$\mathbf{t}-\mathbf{t}^0 = G^0(\mathbf{s}-\mathbf{s}^0). \quad (19)$$

If the call the traveltimes **t** as the vector observed traveltimes $\mathbf{t}^{obs}$, and the slowness **s** and the up-dated slowness $\mathbf{s}^{(k+1)}$, equation (19) can be generalized for the *k*-th iteration:

$$\mathbf{t}^{obs} - \mathbf{t}^{(k)} = G^{(k)}(\mathbf{s}^{(k+1)} - \mathbf{s}^{(k)}),$$

or

$$\Delta\mathbf{t}^{(k)} = G^{(k)}\Delta\mathbf{s}^{(k)}, \quad (20)$$

where $\Delta\mathbf{t}^{(k)}$ corresponds to the misfit between computed and observed traveltimes for a specific model at *k*-th iteration, $G^{(k)}$ contains the elements g_{ji} which corresponds to the distances along the *j*-th raypath across the *i*-th block of model, and $\Delta\mathbf{s}^{(k)}$ corresponds to slowness model update at *k*-th iteration.

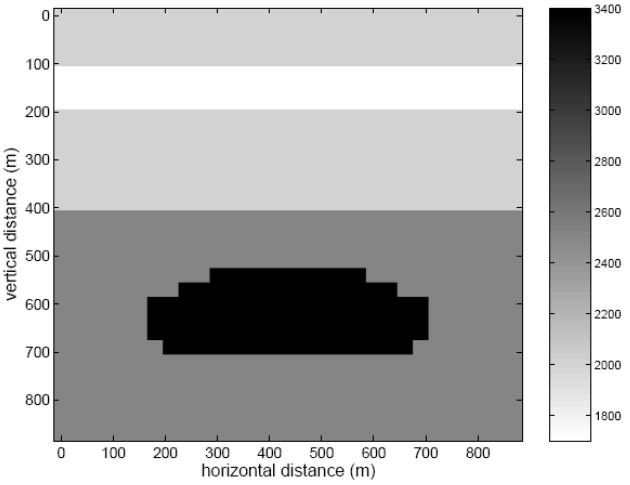
For the linear case, the rays are straight, that is, the ray trajectory does not depend of the slowness distribution. In this case, the equation (20) becomes

$$\mathbf{t} = G\Delta\mathbf{s}. \quad (21)$$

SIMULATIONS

The adopted geological model for simulations can be seen in Figure 5 and it has as main feature a lens simulating a reservoir. The model was discretized into 900 blocks, being each block a square with 30 x 30 meters. The velocity of the homogeneous background medium is 2,000 *m/s*, with a layer that is 100 *m* thick and velocity equals to 1,700 *m/s*. The sound propagation velocity of the target reservoir is 3,400 *m/s*, which is surrounded by a seal with 2,400 *m/s*. The adopted acquisition geometry was well-to-well, with 30 sources and 30 receivers, in such a way that there are 900 rays. Thus the tomographic matrix is square and the system of linear equations is determined with 900 equations and 900 unknowns. The adopted approach can also be applied to solve a system that could also be underdetermined or overdetermined, with rectangular matrices, if needed.

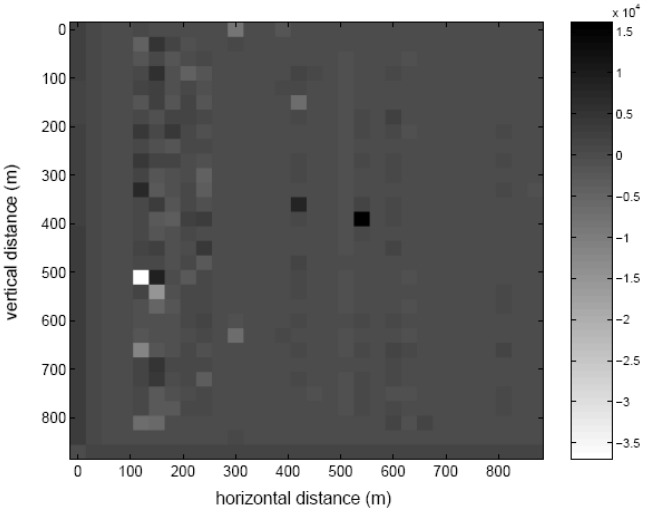
Figure 5 - True model for traveltime tomography simulation. The grayscale color bar indicates the acoustic velocities in *m/s*



In a first simulation we directly used SVD without any regularization technique, in order to deliberately compare with the upcoming inversions using derivative matrices to regularize the solution. The result was not satisfactory, as can be seen in Figure 6, generating an inconsistent and rather unstable estimated model. SVD alone is not enough to stabilize the solution, since the singularity of the matrix greatly disturbs the data. Although being a subjective criterion, a significant improvement can be observed with the truncation of the very small singular values that produce undesirable effects in the matrix inversion. It is

noteworthy that in the next results, using regularization, all singular values were used, that is, the truncation procedure was not applied.

Figure 6- Estimated model with non-truncated SVD, using 900 singular values. The gray color bar indicates de acoustic velocities in *m/s*



The same geological model was used for the simulations with regularization, both with noise free data and with data contaminated with Gaussian noise. In each simulation we calculated the L-curve for three regularization matrices (orders 0, 1 and 2). For each simulation we computed the optimal regularization parameter from the L-curve using three aforementioned methodologies (Hansen's toolbox, visual inspection and the first derivative).

Table 1 summarizes all the results. Due to space limitations, we show the results for first order regularization and the data corrupted with noise factor α = 0.1. Figure 7 shows the L-curve obtained by Hansen's toolbox, which provides the optimal parameter λ^{opt} = 66.87. Using this parameter we obtained the recovered tomogram showed in Figure 8. The RMS error between the true and estimated models was E_m = 4.42 %. We also generated a L-curve that can be seen in Figure 9. By the visual inspection criterion we choose as optimal parameter λ^{opt} = 10,000. Using this parameter we obtained the estimate showed in Figure 10 with E_m = 4.51 %. The derivative of the L-curve in Figure 9 can be seen in Figure 11. The maximum of the curve in Figure 11 corresponds to the parameter λ^{opt} = 100,000, which is marked by a circle. This parameter was used to generate the recovered tomogram shown in Figure 12, with E_m = 4.83 %. Although the regularization parameter provided in Hansen's toolbox was different from the other two methods because of its normalization, the L-curves computed for both methods were similar, as well as the coordinates of the optimal regularization parameter in both L-curves.

Table 1 - Results for the tomographic inversion, comparing the three methods for the extraction of the optimal parameter λ^{opt}, using three levels of noise (α = 0, α = 0.01 and α = 0.1) and three orders of regularization (n = 0, n =1 and n =2). E_m is the RMS relative RMS error between the true model and the estimated model. The last column informs the CPU processing time in seconds

Method 1 - Hansen's toolbox			
α	n	λ^{opt}	E_m (%)
0	0	3.31×10^{-7}	16.79
	1	6.92×10^{-6}	1.10
	2	5.74×10^{-6}	1.14
0.01	0	6.27×10^{-2}	16.84
	1	1.71	3.84
	2	2.69	3.54
0.1	0	5.73×10^{-2}	117.35
	1	66.87	4.42
	2	76.25	4.84
Method 2 - Visual inspection			
0	0	10^{-3}	3.68
	1	10^{-3}	3.46
	2	10^{-3}	2.30
0.01	0	10^{-2}	3.29
	1	10.00	3.80
	2	1.00	3.23
0.1	0	1.00	16.88
	1	10^{4}	4.51
	2	10^{3}	4.91
Method 3 - First derivative			
0	0	10^{-6}	2.39
	1	10^{-6}	1.11
	2	10^{-6}	1.08
0.01	0	10.00	3.71
	1	10^{2}	3.90
	2	1.00	3.23
0.1	0	10^{2}	5.30
	1	10^{5}	4.83
	2	10^{6}	5.28

Figure 7 - L-curve from Hansen's toolbox, first order regularization, λ^{opt} = 66.87

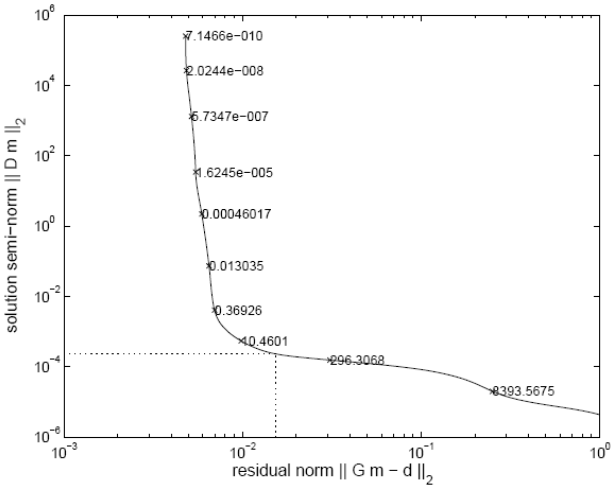


Figure 8 - Estimated model using λ^{opt} = 66.87 from Hansen's toolbox. α = 0.1, E_m = 4.42 %. The gray color bar indicates the acoustic velocities in *m/s*

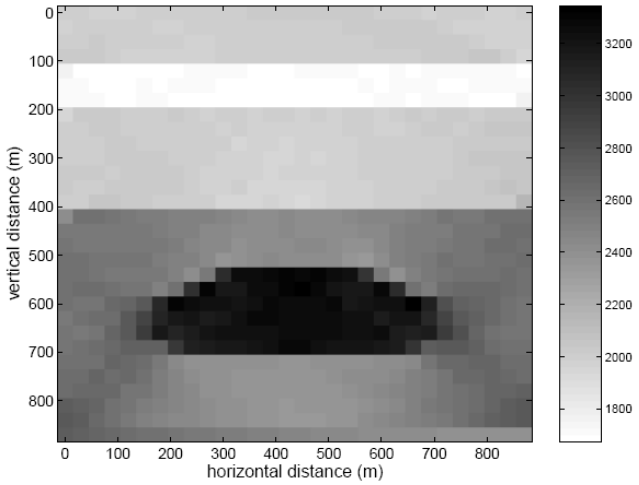


Figure 9 - L-curve for visual inspection and calculation of the first derivative, first order regularization

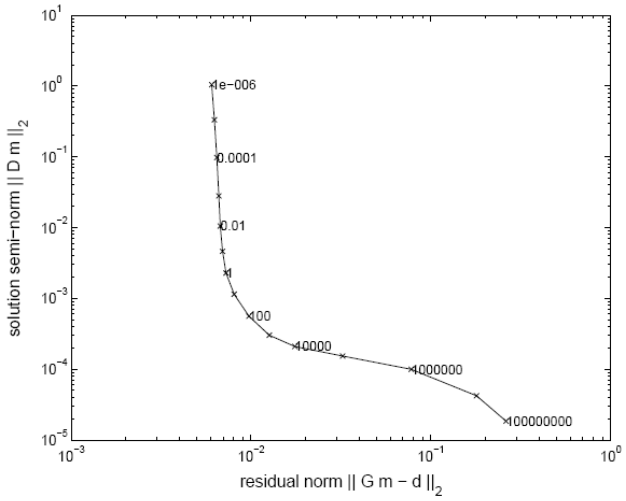


Figure 10 - Estimated model using the visual inspection of the L-curve, first order regularization, λ^{opt} = 10,000, α = 0.1, E_m = 4.51 %. The gray color bar indicates de acoustic velocities in *m/s*

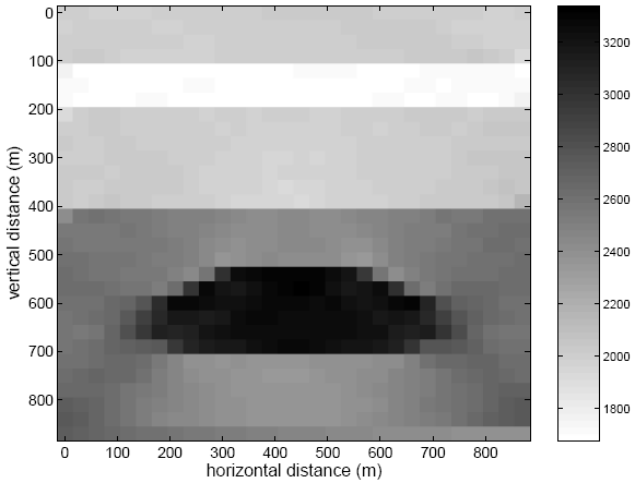


Figure 11 - First derivative of the L-curve, first order regularization, where is possible to see λ^{opt} = 100,000 with a circle

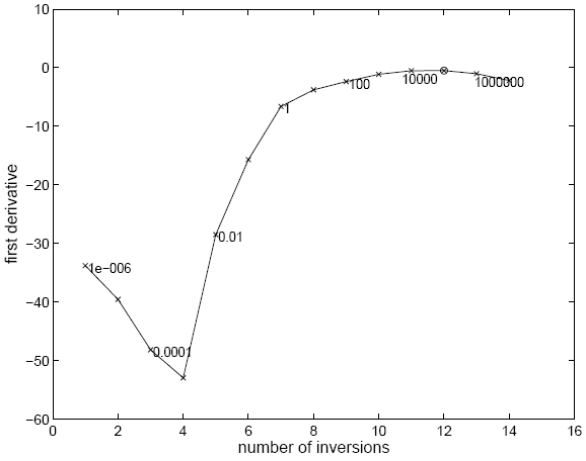
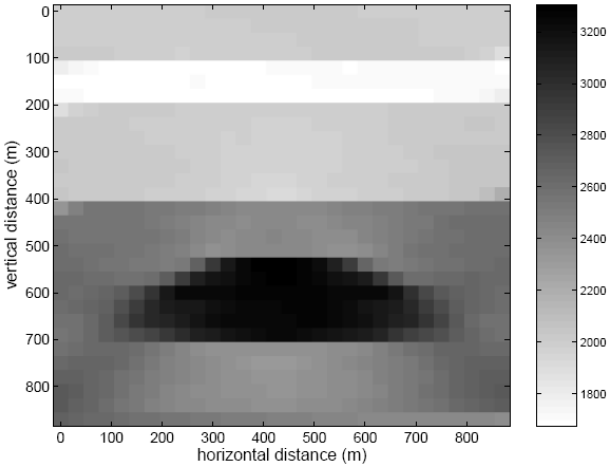


Figure 12 - Estimated model using the first derivative of the L-curve, first order regularization, λ^{opt} = 100,000, α = 0.1, E_m = 4.83 %. The gray color bar indicates de acoustic velocities in *m/s*



From the qualitative point of view, Figures 8, 10 and 12 show great resemblance. From a quantitative standpoint, the RMS error between the true model and the estimated model, associated to Figures 8, 10 and 12 are very close, the first being slightly the best approach. However, when analyzing Table 1 as a whole, which summarizes the results of 27 simulations, we found that approach 3 is more robust, since the RMS error between real and recovered models is always less than 5%. Besides, the approaches 2 and 3 are considerably superior in the case of zero order regularization.

CONCLUSIONS

L-curve has an important role in the solution of ill-posed inverse problems that use the regularization technique. This is because the L-curve is one of the best ways to determine the optimal regularization parameter. Although the use of L-curve represents an additional computation cost, it has the advantage to skip the almost always non-consistent method of trial and error for the choice of the optimum λ. Despite the fact that the numerical treatment is different, the L-curve obtained from Hansen's toolbox (method 1), which uses GSVD, is similar to that calculated using FORTRAN for the methods 2 and 3, which uses standard SVD. Thus the estimated models are in general very similar, although the optimum λ is very different.

Hansen's toolbox provides the optimum λ in an automatic way, and its computational time is lower if compared to the other two approaches. GSVD decomposes the two matrices just once, while in our approach we perform an inverse procedure by SVD for each λ. The simulations in traveltime tomography, with several noise levels, and three regularization orders, allowed us to validate the three approaches for the extraction of optimum λ from the L-curve. With the increase of additive noise, one must also increase the amount of regularization, as can be seen in Table 1.

The second and the third criteria provided the best result for zero order regularization. The first order regularization was the most effective in all the three methodologies. In general, the third criterion provided the best results. Despite the fact that Hansen's toolbox is the fastest method, sometimes the selected λ was not the best one. Besides, the available toolbox cannot be applied to underdetermined problems. On the other hand, our FORTRAN code for SVD computation does not have such limitation. Moreover, in cases where the L-curve does not have the shape of the letter L, we found that the method of visual inspection, together with the calculation of the first derivative of the curve becomes a better option.

The estimated models in traveltime tomography were quite satisfactory. The true model, which simulates a realistic case of an oil reservoir where the area of interest was divided into a mesh of 30 x 30 blocks, was well recovered even in the presence of noise. One can also notice that the modeling using straight ray tracing is feasible where there are no abrupt transitions in velocity.

ACKNOWLEDGMENTS

F. A. Terra would like to thank the PIBIC/UFBA program, as well as FAPESB and CNPq, for a two-year scientific initiation scholarship. A. Bassrei would like to thank CNPq for the projects 307.427/2010-2 and 308.690/2013-3 (research fellowship), CNPq and PETROBRAS for the support to Instituto Nacional de Ciência e Tecnologia de Geofísica de Petróleo (INCT-GP), and FINEP for the support to the Rede Cooperativa de Pesquisa em Geofísica de Exploração (Rede 01).

REFERENCES

BASSREI, A.; RODI, W. L. Regularization and inversion of linear geophysical data. In: INTERNATIONAL CONGRESS OF THE BRAZILIAN GEOPHYSICAL SOCIETY, 3, Rio de Janeiro, Brazil, 1993. V. 1, p. 111-116, 1993.

FARQUHARSON, C. G.; OLDENBURG, D. W. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. **Geophysical Journal International**, v. 156, p. 411-425, 2004.

HANSEN, P. C. Analysis of discrete ill-posed problems by means of the L-curve. **Society for Industrial and Applied Mathematics Review**, v. 34, p. 561-580, 1994.

HANSEN, P. C. **Rank-Deficient and Discrete Ill-Posed Problems.** Philadelphia: Society for Industrial and Applied Mathematics, 1998.

LANCZOS, C. **Linear Differential Operators**. London: Van Nostrand, 1961.

LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. **Quarterly of Applied Mathematics**, v. 2, p.164—168, 1944.

MARQUARDT, D. W. An algorithm for least-squares of nonlinear parameters. **Journal of Society of Industrial and Applied Mathematics**, v. 11, p. 431-441, 1963.

RAY, R. D.; SANCHEZ, B. V. Improved smoothing of an altimetric ocean-tide model with global Proudman functions. **Geophysical Journal International**, v. 118, p. 788-794, 1994.

SANTOS, E. T. F.; BASSREI, A.; COSTA, J. Evaluation of L-curve and Θ-curve approaches for the selection of regularization parameter in anisotropic traveltime tomography. **Journal of Seismic Exploration**, v. 15, p. 245-272, 2006.

SANTOS, E. T. F.; BASSREI, A. L- and Θ-curve approaches for the selection of regularization parameter in geophysical diffraction tomography. **Computers & Geosciences**, v. 33, p. 618-629, 2007.

TITTERINGTON, D. M., General structure of regularization procedures in image reconstruction. **Astronomy and Astrophysics**, v. 144, p. 381-387, 1985.

TWOMEY, S. On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. **Journal of the Association of Computing Machines,** v. 10, p. 97-101, 1963.

YAO, Z. S.; ROBERTS, R. G. A practical regularization for seismic tomography. **Geophysical Journal International**, v. 138, p. 293-299, 1999.