

## Conjugate Gradient Method for the Solution of Inverse Problems: Application in Linear Seismic Tomography

T.E.B. BRUFATI<sup>1</sup>, S.P. OLIVEIRA<sup>1\*</sup> and A. BASSREI<sup>2</sup>

Received on May 5, 2014 / Accepted on August 25, 2015

**ABSTRACT.** We consider the conjugate gradient method for the normal equations in the solution of discrete ill-posed problems arising from seismic tomography. We use a linear approach of travelttime tomography that is characterized by an ill-conditioned linear system whose unknowns are the slownesses in each block of the computational domain. The algorithms considered in this work regularize the linear system by stopping the conjugate gradient method in an early iteration. They do not depend on the singular-value decomposition and represent an attractive and economic alternative for large-scale problems. We review two recently proposed stopping criteria and propose a modified stopping criterion that takes into account the oscillations in the approximate solution.

**Keywords:** seismic tomography, conjugate gradient method, truncated iteration.

### 1 INTRODUCTION

Exploration seismology, or seismics, is the field of geophysics that is most employed for subsurface imaging in the oil industry, and uses an ensemble of techniques based upon the theory of propagation of elastic and acoustic waves. One of such techniques is seismic tomography, that became an important tool in reservoir geophysics due to its high resolution.

Tomographic reconstruction is a special kind of inversion procedure that allows one to estimate material properties from their line integrals. Tomography first appeared in medicine for imaging the human body. A set of sources and receivers (usually of X-rays) rotates around the patient in order to get a complete scan, providing images of high resolution. Unlike medicine, in geophysics one cannot make a complete turn around the object of study. And rather than using X-rays, one uses electromagnetic waves with lower frequency, as well as mechanical waves, which is the object of the present work.

There are two general classes of seismic tomography: waveform and travelttime tomography, which are dynamical and kinematic approaches, respectively. Since the 1980s, these techniques

---

\*Corresponding author: Saulo Pomponet Oliveira.

<sup>1</sup>Departamento de Matemática, Universidade Federal do Paraná, 81531-980 Curitiba, PR, Brasil.  
E-mails: tuanny@redispapneus.com.br; saulopo@ufpr.br

<sup>2</sup>CPGG/IGEO and INCT-GP, Universidade Federal da Bahia, 40170-115 Salvador, BA, Brasil. E-mail: bassrei@ufba.br

have been adopted in exploration seismology as a data inversion method which is known for its high resolution, when compared with conventional seismic acquisition geometries. We will focus on traveltimes tomography, which uses the traveltimes between sources and receivers. These traveltimes constitute the vector of parameters  $\mathbf{t}$ , which is the input data in tomographic inversion. The matrix  $G$  used in tomography describes the geometry of the rays that connect sources and receivers. The output data of the tomographic inversion is the vector of model parameters,  $\mathbf{s}$ , which contains the slowness (the reciprocal of the velocity) field, in the isotropic case, or density normalized elastic parameters, in the anisotropic case.

One of the most popular approaches to find the least-squares solution to the linear system  $G\mathbf{s} = \mathbf{t}$  is the singular-value decomposition (SVD) [9]. However, since the matrix  $G$  is usually sparse, iterative methods are the usual choices for large-scale problems [8, 9, 16]. Taking into account that several authors reported that iterative methods based on Krylov subspaces, such as the conjugate gradient method for the normal equations and the LSQR method [10], performed better than stationary iterative methods such as SIRT and ART techniques [9, 7, 19], we focus on conjugate gradient methods in this work.

The difficulties on inversion arise from the fact that it typically involves ill-conditioned matrices, in order that the solutions are highly sensitive to data perturbations, which characterizes an ill-posed problem. Other sources of difficulties are the presence of noise, scarce information and an inaccurate model discretization. Perhaps the most popular procedure to alleviate these drawbacks is to resort to a regularization technique by means of derivative matrices, which is known in the literature as Tikhonov regularization. This technique employs a regularization parameter  $\lambda$  whose choice is crucial to the correctness of the solution. There are several classical techniques available for selecting the regularization parameter, such as the L-curve and the generalized cross-validation (GCV) method (see, e.g., [6]). Some examples of alternative techniques are the  $\Theta$ -curve [11, 12] and the fixed-point iteration [2].

One common alternative to Tikhonov regularization for iterative methods is to break the iterations at an early stage [5, 13], which requires a more strict stopping criterion. In this work, we consider two stopping criteria. The first criterion is based on an estimate of the GCV functional that requires a small computational cost [15]. The second criterion, which has an even lower cost, depends on the product of the norm of the residual and the norm of the current iterate [3, Sec. 4.1].

The remainder of this paper is organized as follows. In the next section, we review the mathematical model of linearized tomography and consider the simpler case of linear tomography, where the inverse problem reduces to the solution of a linear system. In Section 3, we present the conjugate gradient algorithm with truncated iteration and, motivated by the work of Van-De-car and Snieder [20], who proposed a penalization of the first and second derivatives of the estimated model parameters, we modify the minimal product criterion from [3] in order to take into account the oscillations of the current iterate. These criteria are compared by means of numerical experiments in Section 4 and further discussed in Section 5.

## 2 LINEARIZED SEISMIC TOMOGRAPHY

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

$$t = \int_{r(s)} s(x, z) dl, \quad (2.1)$$

where  $r$  is the path of the trajectory along which the integration is performed,  $dl$  is the ray element, and  $s(x, z)$  is the slowness of the medium at the point  $(x, z)$ , where  $x$  and  $z$  represent the horizontal and vertical coordinates, respectively.

The slowness model can be parameterized using a regular grid of  $N$  blocks,

$$s(x, z) = \sum_{j=1}^N s_j B_j(x, z), \quad (2.2)$$

where the basis function  $B_j(x, z)$  is the characteristic function, which is 1 on the  $j$ -th block and zero otherwise. This model representation leads to the relation

$$\mathbf{t} \equiv \mathbf{g}(\mathbf{s}), \quad \mathbf{g}(\mathbf{s}) \equiv \sum_{j=1}^N s_j \int_{r(s)} B_j(x, z) dl, \quad (2.3)$$

which is nonlinear, since the ray path depends on the slowness. The vector  $\mathbf{t} \in \mathbb{R}^M$  represents the observed travel times of each ray. The vector  $\mathbf{s} \in \mathbb{R}^N$  represents the slowness in each block. We consider a first-order truncated Taylor series approximation around  $\mathbf{s}^{(0)}$ :

$$\mathbf{t} = \mathbf{t}^{(0)} + \frac{\partial \mathbf{g}}{\partial \mathbf{s}} (\mathbf{s} - \mathbf{s}^{(0)}), \quad \mathbf{t}^{(0)} = \mathbf{g}(\mathbf{s}^{(0)}), \quad (2.4)$$

and carry out an iterative process where, at each step  $k$ , we have the following linear relation between the the vector  $\mathbf{t}^{(k)}$  of traveltimes of all rays and the vector  $\mathbf{s}^{(k)}$  with the current estimated slowness field:

$$\mathbf{t}^{(k)} = G^{(k)} \mathbf{s}^{(k)}. \quad (2.5)$$

The matrix  $G^{(k)}$  contains elements  $g_{ij}$  that correspond to the distances that the  $i$ -th ray travels in the  $j$ -th block. This matrix can be constructed with a ray tracing algorithm [1]<sup>3</sup>. Let us denote by  $\mathbf{t}^{obs}$  the observed traveltime. By applying the expansion (2.4) on each component of  $\mathbf{t}^{obs}$  and using approximation (2.5), we find

$$\Delta \mathbf{t}^{(k)} = G^{(k)} \Delta \mathbf{s}^{(k)}, \quad (2.6)$$

where  $\Delta \mathbf{t}^{(k)} = \mathbf{t}^{obs} - \mathbf{t}^{(k)}$  and  $\Delta \mathbf{s}^{(k)} = \mathbf{s}^{(k+1)} - \mathbf{s}^{(k)}$ .

We want to find the vector  $\Delta \mathbf{s}^{(k)}$  that minimizes  $\| G^{(k)} \Delta \mathbf{s}^{(k)} - \Delta \mathbf{t}^{(k)} \|_2$ . This is called a least squares problem, which can be formally stated as finding the minimum of the following objective function:

$$\Phi(\Delta \mathbf{s}^{(k)}) = (\Delta \mathbf{t}^{(k)} - G^{(k)} \Delta \mathbf{s}^{(k)})^T (\Delta \mathbf{t}^{(k)} - G^{(k)} \Delta \mathbf{s}^{(k)}). \quad (2.7)$$

<sup>3</sup>A FORTRAN ray tracing program for this purpose is available in [17].

The estimated solution, also called least squares solution, is

$$[G^{(k)}]^T G^{(k)} \Delta s^{(k),est} = [G^{(k)}]^T \Delta t^{(k)}. \tag{2.8}$$

Least-squares solutions very often do not provide good results and sometimes they do not even exist. In order to solve this problem we use a tool of regularization or smoothing: the ill-conditioning of the matrix  $G^{(k)}$  is regularized and the unstable least-squares estimate  $\Delta s^{(k),est}$  is consequently smoothed to greatly reduce the possibility of wild noise-induced fluctuation in  $\Delta t^{(k)}$ , hopefully without distorting the resulting smoothed image too far from the true solution  $\Delta s^{(k)}$  [18].

In this work, we have considered linear tomography, where the geometry of the rays does not depend on the distribution of velocities [7]. In other words, the rays are straight, and there is no need to do further ray tracing, nor to update the vector of estimated slowness. In general, this approximation is valid only for small velocity (or slowness) contrasts.

Thus, eq. (2.8) now becomes

$$G^T G s = G^T t, \tag{2.9}$$

where the solution  $s = s^{est}$  will be obtained through conjugate gradient schemes as presented in the next section.

### 3 CONJUGATE GRADIENT METHOD AND STOPPING CRITERIA

We consider the classical conjugate gradient method for normal equations to solve the least-squares linear system (2.9), which is often referred to as CGLS [3, 10, 15]. Given a initial guess  $s_0$  and a maximum number of iterations  $k_{max}$ ,

```

r0 ← t - Gs0, z0 ← G^T r0, and p0 ← z0;
repeat
  wk ← Gpk;
  αk ← (zk^T zk)/(wk^T wk);
  sk+1 ← sk + αkpk;
  rk+1 ← rk - αk wk;
  zk+1 ← G^T rk+1
  βk ← (zk+1^T zk+1)/(zk^T zk);
  pk+1 ← zk+1 + βkpk;
until Φ(k) reaches a local minimum or k > kmax
    
```

The matrix  $G^T G$  is not explicitly computed, since it usually has a higher number of nonzero entries and a higher condition number than  $G$  [16].

The stopping criterion is driven by the function  $\Phi(k)$ . We consider the following approaches:

- Monte-Carlo GCV [15]:

Given  $w \sim N(\mathbf{0}, I_{M \times M}) \in \mathbb{R}^M, \delta = 10^{-4}$ ,

$$\Phi(k) = V_{NL}(k) = \frac{\|t - Gs^k\|^2}{M\Phi_{NL}(k)}, \tag{3.1}$$

$$\Phi_{NL}(k) = \left( \frac{\mathbf{w}^T [\mathbf{w} - G(s_+^k - s_-^k)/(2\delta)]}{\mathbf{w}^T \mathbf{w}} \right)^2, \quad (3.2)$$

where  $s_{\pm}^k$  are conjugate gradient iterates corresponding to the right-hand side  $\mathbf{t} \pm \delta \mathbf{w}$ .

- Minimal product [3]:

$$\Phi(k) = \psi(k) = \|\mathbf{t} - G\mathbf{s}^k\| \|\mathbf{s}^k\|. \quad (3.3)$$

Recalling the penalization of the first and second derivatives of the estimated model parameters proposed in [20], we modify the minimal product criterion from [3] by introducing an additional term that penalizes large oscillations in the current iterate:

$$\Phi(k) = \psi(k) = \|\mathbf{t} - G\mathbf{s}^k\| \sqrt{\|\mathbf{s}^k\|^2 + \|D\mathbf{s}^k\|^2}, \quad (3.4)$$

where  $D$  is the first-order finite difference operator given in matrix form as

$$D = \frac{1}{h} \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \quad h = \frac{1}{N}. \quad (3.5)$$

#### 4 NUMERICAL RESULTS

We consider the model shown in Figure 1 (left), which is described as a non symmetric anticlinal with tectonic origin. Such situation has great relevance in oil exploration, since it has structural traps with folds that could accumulate hydrocarbons. The reservoir is represented by a porous sandstone (2,100 m/s) and the sealing by a impermeable shale (2,500 m/s). The model was discretized in  $N = 800$  blocks, where each block is a  $10 \text{ m} \times 10 \text{ m}$  square.

For the simulations we considered 31 sources in one borehole and 31 receivers in the other, in such a way that we have  $M = 961$  rays (Fig. 1, right). The observed data  $\mathbf{t}^{obs}$  was perturbed with the addition of a uniformly distributed random noise with amplitude  $\alpha$ , with  $\alpha = 0, 10^{-4}, 10^{-3}$ , and  $10^{-2}$ . We denote these perturbed vectors as  $\mathbf{t}^\alpha$ .

In the following we evaluate the following relative mean square errors: the error of the perturbed data with respect to the observed data, the calculated traveltimes with respect to the perturbed data, and the estimated velocities and slownesses with respect to the true model. These errors are defined as follows:

$$\begin{aligned} \epsilon_\alpha &= \frac{\|\mathbf{t}^\alpha - \mathbf{t}^{obs}\|_2}{\|\mathbf{t}^{obs}\|_2}, & \epsilon_t &= \frac{\|\mathbf{t}^{calc} - \mathbf{t}^\alpha\|_2}{\|\mathbf{t}^\alpha\|_2} \quad (\mathbf{t}^{calc} = G\mathbf{s}^{est}), \\ \epsilon_v &= \frac{\|\mathbf{v}^{est} - \mathbf{v}^{true}\|_2}{\|\mathbf{v}^{true}\|_2}, & \epsilon_s &= \frac{\|\mathbf{s}^{est} - \mathbf{s}^{true}\|_2}{\|\mathbf{s}^{true}\|_2}. \end{aligned} \quad (4.1)$$

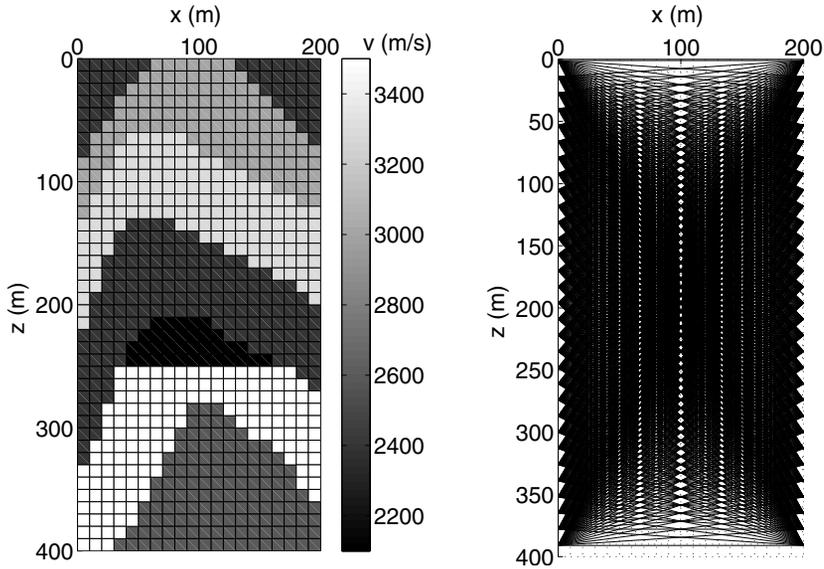


Figure 1: Synthetic model, true velocities (left) and ray tracing diagram (right).

Figures 2-3 compare the error of the  $k$ -th iterate with the functions  $\Phi(k)$  for each approach considered in previous section. We employed  $k_{max} = N$ . In analogy with [15], the error is normalized so that its scale is comparable with the scale of the stopping criterion functions. The minimal product (MP) function is the most economical criterion, but did not predict early enough the error increase when the perturbations were given by  $\alpha = 10^{-3}$  and  $\alpha = 10^{-2}$ . On the other hand, the modified minimal product (MMP) function, whose computational complexity is similar to MP, had similar results to the Monte-Carlo CGV (MC-GCV).

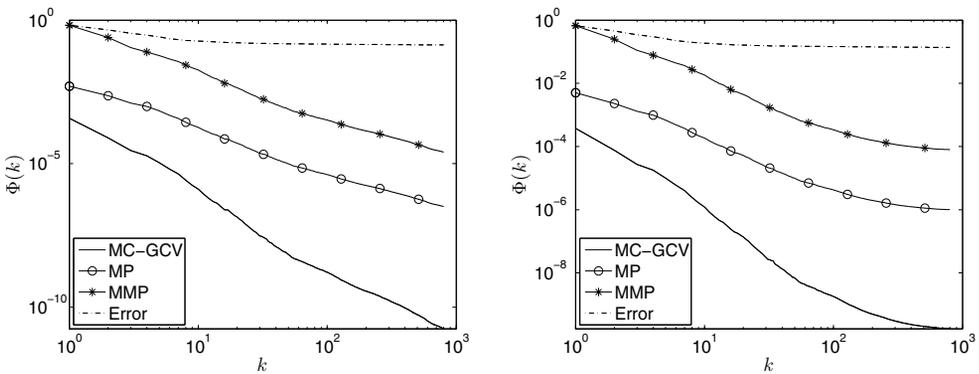


Figure 2: Stopping criteria for  $\alpha = 0$  and  $\alpha = 10^{-4}$ .

The iterative scheme (2.9) with the MMP criterion yields the reconstructed models shown in Figures 4-5. Table 1 shows the number of iterations and the mean square errors of each inversion.

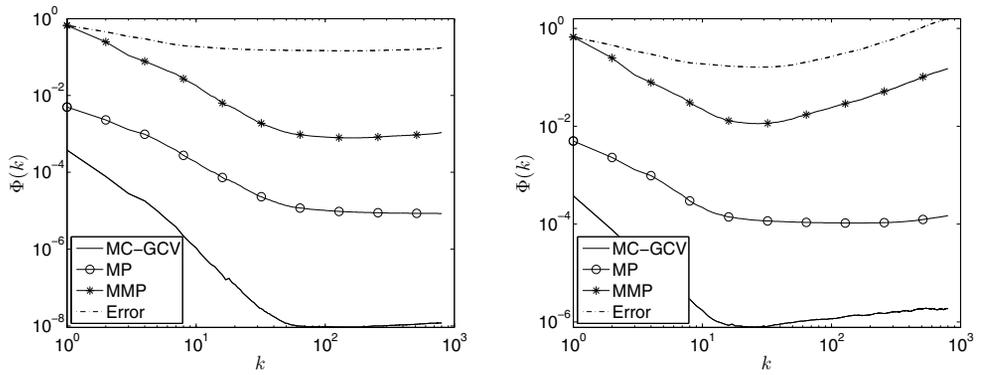


Figure 3: Stopping criteria for  $\alpha = 10^{-3}$  and  $\alpha = 10^{-2}$ .

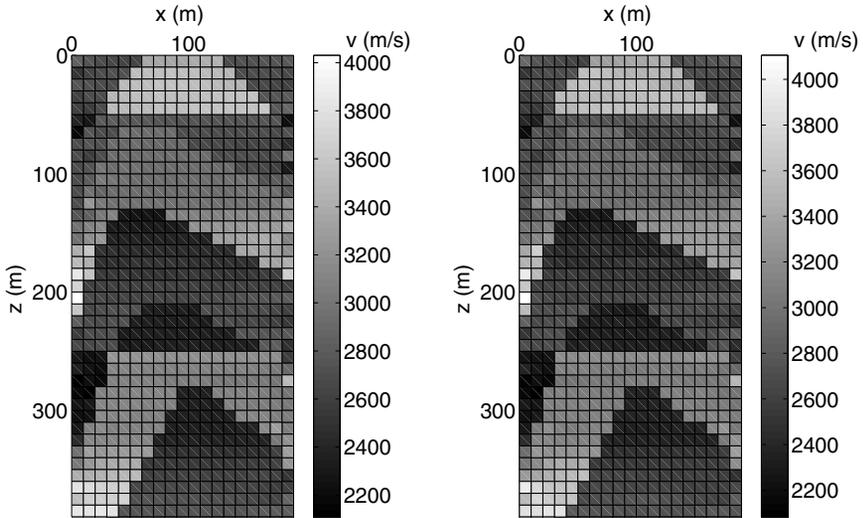


Figure 4: Reconstructed models with  $\alpha = 0$  (left) and  $\alpha = 10^{-4}$  (right).

Table 1: Number of iterations and mean square errors (in percentage).

$\alpha$	$k$	$\epsilon_\alpha$	$\epsilon_t$	$\epsilon_v$	$\epsilon_s$
0	800	0	0.001093	10.6299	10.095
$10^{-4}$	800	0.005749	0.003459	10.7056	10.126
$10^{-3}$	161	0.056209	0.031956	11.5065	10.778
$10^{-2}$	28	0.569306	0.408862	12.5892	11.995

The estimated models are satisfactory, allowing the identification (location and velocity) of the different layers, including the target reservoir. However, for the simulation with the highest noise level ( $\alpha = 10^{-2}$ ), the layered geometry is not well delineated.

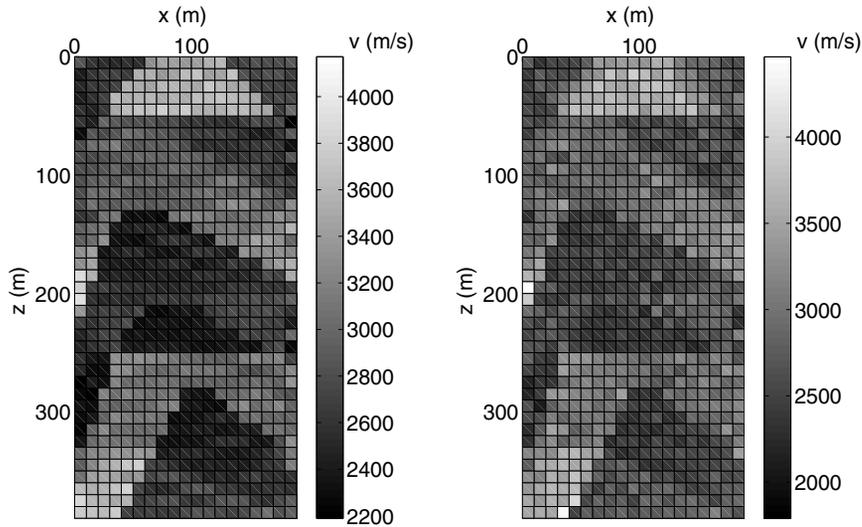


Figure 5: Reconstructed models with  $\alpha = 10^{-3}$  (left) and  $\alpha = 10^{-2}$  (right).

## 5 CONCLUSIONS

We considered the solution of an ill-conditioned least-squares problem arising from the discretization of a model for linear seismic tomography, which is a tool with wide application in reservoir geophysics. The discrete problem was presented as a system of linear equations, which we solved by the standard conjugate gradient method regularized by stopping in an early stage.

In the numerical experiments with a synthetic geological model, we perturbed the observed data with different levels of noise. The estimated model parameters were all satisfactory, although the resolution in the reconstructed velocity tomograms were compromised when the noise level was higher.

We expect that the effectiveness of the stopping criteria would be better in the absence of noise (or for low noise) if an appropriate preconditioning of the linear system is employed. We have also performed numerical experiments with conjugate gradient methods with Tikhonov regularization and noticed that, also in this case, a large number of iterations is needed for convergence. Some of the preconditioners proposed in the literature that might be appropriate are based on the Fast Fourier Transform [4] and on the Algebraic Reconstruction Technique [14]. Moreover, the preliminary results presented herein do not provide sufficient evidence that the MMP criterion is a robust alternative to the MP and the MC-GCV criteria. This motivates further testing and an error analysis in future works.

## ACKNOWLEDGMENTS

The authors would like to thank CNPq for sponsoring the National Institute of Science and Technology in Petroleum Geophysics (INCT-GP), FINEP for sponsoring the CTPETRO Network in Exploration Geophysics (Rede 01) and Naiane Pereira de Oliveira for providing the synthetic model employed in this work. T. Brufati would like to thank UFPR for the TN scholarship (2011-2012). S.P. Oliveira and A. Bassrei would like to thank CNPq for the research fellowships 306083/2014-0 and 308690/2013-3, respectively.

**RESUMO.** Consideramos o uso do método do gradientes conjugados para as equações normais na solução de problemas inversos decorrentes da tomografia sísmica. Empregamos um modelo linear de tomografia de tempos de trânsito, que é caracterizado por um sistema linear mal condicionado em que as incógnitas representam a vagarosidade em cada bloco do domínio computacional. Os algoritmos estudados neste trabalho regularizam o sistema linear interrompendo precocemente as iterações do método do gradientes conjugados. Estes algoritmos não dependem da decomposição em valores singulares, representando uma alternativa atraente para problemas de grande porte. Revisamos dois critérios de parada recentes e desenvolvemos um critério de parada modificado que penaliza oscilações na solução aproximada.

**Palavras-chave:** tomografia sísmica, método de gradientes conjugados, iteração truncada.

## REFERENCES

- [1] A.H. Andersen & A.C. Kak. Digital ray tracing in two-dimensional refractive fields. *Journal of the Acoustic Society of America*, **72**(5) (1982), 1593–1606.
- [2] F.S.V. Bazán. Fixed-point iterations in determining the Tikhonov regularization parameter. *Inverse Problems*, **24**(3) (2008), 035001.
- [3] F.S.V. Bazán, M.C.C. Cunha & L.S. Borges. Extension of GKB-FP algorithm to large-scale general-form Tikhonov regularization. *Numerical Linear Algebra with Applications*, **21**(3) (2014), 316–339.
- [4] R. Chan, J. Nagy & R. Plemmons. FFT-based preconditioners for Toeplitz-block least squares problems. *SIAM Journal on Numerical Analysis*, **30**(6) (1993), 1740–1768.
- [5] H. Fleming. Equivalence of regularization and truncated iteration in the solution of ill-posed image reconstruction problems. *Linear Algebra and its Applications*, **130** (1990), 133–150.
- [6] P.C. Hansen. *Rank-deficient and discrete ill-posed problems*. SIAM, Philadelphia, (1998).
- [7] S. Ivansson. Seismic borehole tomography – theory and computational methods. *Proceedings of the IEEE*, **74**(2) (1986), 328–338.
- [8] E.-J. Lee, H. Huang, J. Dennis, P. Chen & L. Wang. An optimized parallel LSQR algorithm for seismic tomography. *Computers and Geosciences*, **61** (2013), 184–197.
- [9] G. Nolet. Solving or resolving inadequate and noisy tomographic systems. *Journal of Computational Physics*, **61**(3) (1985), 463–482.

- [10] C. Paige & M. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, **8**(1) (1982), 43–71.
- [11] E. Santos & A. Bassrei. L- and  $\Theta$ -curve approaches for the selection of regularization parameter in geophysical diffraction tomography. *Computers and Geosciences*, **33**(5) (2007), 618–629.
- [12] E.T.F. Santos, A. Bassrei & J. Costa. Evaluation of L-curve and  $\Theta$ -curve approaches for the selection of regularization parameter in anisotropic travelttime tomography. *Journal of Seismic Exploration*, **15** (2006), 245–272.
- [13] R.J. Santos. Equivalence of regularization and truncated iteration for general ill-posed problems. *Linear Algebra and its Applications*, **236** (1996), 25–33.
- [14] R.J. Santos. Preconditioning conjugate gradient with symmetric algebraic reconstruction technique (ART) in computerized tomography. *Applied Numerical Mathematics*, **47**(2) (2003), 255–263.
- [15] R.J. Santos & A. de Pierro. The effect of the nonlinearity on GCV applied to conjugate gradients in computerized tomography. *Computational and Applied Mathematics*, **25** (2006), 111–128.
- [16] J. Scales. Tomographic inversion via the conjugate gradient method. *Geophysics*, **52**(2) (1987), 179–185.
- [17] H.A. Schots. Well-to-well and well-to-surface seismic tomography using direct waves (in Portuguese). M.Sc. Dissertation, Universidade Federal da Bahia, (1990).
- [18] D.M. Titterton. General structure of regularization procedures in image reconstruction. *Astronomy and Astrophysics*, **144** (1985), 381–387.
- [19] A. van der Sluis & H. van der Vorst. SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems. *Linear Algebra and its Applications*, **130** (1990), 257–303.
- [20] J. VanDecar & R. Snieder. Obtaining smooth solutions to large, linear, inverse problems. *Geophysics*, **59**(5) (1994), 818–829.