Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.98–110.

AN INTRODUCTION TO ROBUST TOMOGRAPHY METHODS

UTKU KALE–MIHÁLY DOBRÓKA*

Department of Geophysics, University of Miskolc, Hungary
**dobroka@uni-miskolc.hu*

Abstract: Seismic tomography is an important imaging tool in mining, and oil & gas exploration. The aim of this article is to test outlier sensitivity for a medium sized tomography example applied to synthetic data. Throughout the test, tomographic algorithms were applied to demonstrate the powers of algorithms. One of the major problems in seismic tomography is to find an effective tomography algorithm that gives acceptable result just not only for Gaussian data but also in the case of non-Gaussian distribution of noise, for example when a small portion of data contains large error. The purpose of test was to reduce the sensitivity to outliers by using most common algorithms in use today: Conjugated Gradients (CG) and Simultaneous Iterative Reconstruction Technique (SIRT). We also modified the CG and SIRT by utilizing Cauchy-Steiner Weights to define a robust tomography algorithm.

***Keywords**: Robust tomography, Conjugated Gradients (CG), Simultaneous Iterative Reconstruction Technique (SIRT), Model Distance, Data Distance, Cauchy-Steiner Weights, ART Correction, Gaussian and non-Gaussian data.*

1. INTRODUCTION

In this paper, robust tomography images are produced by utilizing some generalized tomography algorithms. Seismic tomography method plays an important role to solve some geophysical problems in mining as well as in oil and gas exploration. In practice, measured data always contain noise that can cause considerable distortion on the tomographic reconstruction. Non-Gaussian noise distribution (including outliers) is extremely dangerous in distorting the tomographic reconstruction. In order to reduce the outlier sensitivity, we have applied a generalized version of some tomography algorithms. Traditionally, least squares problems in tomography have been solved by row action methods such as Algebraic Reconstruction Technique (ART) or Simultaneous Iterative Reconstruction Technique (SIRT). It was also proved [1] that Conjugate Gradient (CG) method can also be used in even large-scale tomographic least squares inversion. It is well known in inverse problem theory that the traditional least square methods give optimal results only if the data noises follow a Gaussian distribution. However, in practice Gaussian distribution seldom occurs. This means that the traditional CG SIRT approaches will not give robust results for non-Gaussian data. Cauchy inversion is frequently used in geophysical inversion as a robust optimization technique [2]. In this paper we modified the traditional algorithms by utilizing the Cauchy-Steiner Weights in order to make the CG and SIRT methods more robust.

2. EXPERIMENTAL

Using synthetic travel-time data the generalized (robust) tomography algorithms are tested in a medium-sized tomography example. In experiments, the tomography model was defined by a rectangular test area of the size 100 × 100 cells *(Figure 1)*. In the model three anomalies with velocities 5 km/s (marked red) are located in a homogeneous background of 4 km/s velocity (blue). Sources and receivers were positioned along the *x* and *y* axes in an arrangement fulfilling the requirement of full tomographic ray coverage, so the theoretical (noiseless) travel time data were computed along 60,000 ray traces. In the model, there are 60,000 known equations and 10,000 unknown parameters, so this is an overdetermined inverse problem.

Two datasets were generated for the tests. In Dataset I the ideal (noiseless) dataset was contaminated with Gaussian noise (the size of the noise is 1% of the theoretical travel-time). Thus, Dataset I follows the Gaussian distribution. In order to simulate non-Gaussian noise, we produce another dataset that is highly distorted. The second dataset was created from the first one by adding 20% extra noise to a randomly selected 20% portion of the data. That means 80% of Dataset II contains 1% noise parameters and 20% of Dataset II contains outliers (20% extra noise).

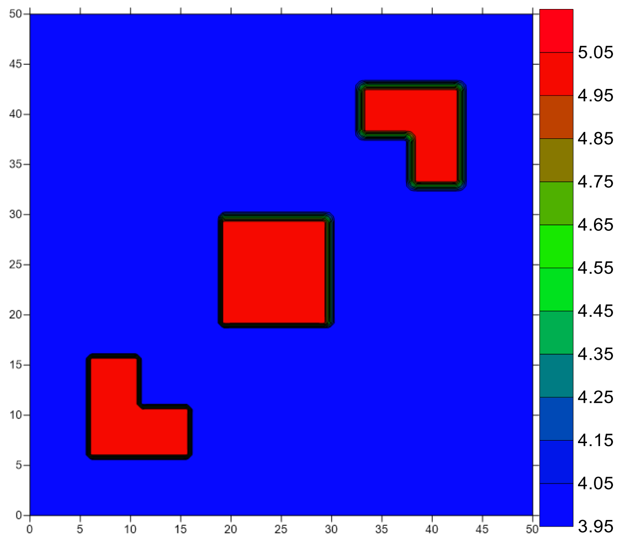

Figure 1. *The model used for numerical tests*

Dataset II is very noisy, and thus not a typical dataset in practice. However, to study the tomographic algorithms it is better to use a challenging dataset to demonstrate the power of the method. (In all of our figures the velocity is measured in km/s units, the distances are assumed in units of cell's size along x and y directions, respectively.)

3. Model distance and Data distance

In order to characterize the accuracy of the reconstruction the (relative) model distance

$$D_{mod} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\frac{x_j - x_j^{(0)}}{x_j^{(0)}}\right)^2} \tag{1}$$

was used. Here x_j and $x_j^{(0)}$ denotes the slownesses (reciprocal values of the propagation velocities) in the *j*-th cell of the reconstructed image and the model, respectively, and *M* is the number of cells. Similarly, the relative data distance was also calculated:

$$D_{dat} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{t_k^{(input)} - t_k^{(reconstructed)}}{t_k^{(input)}}\right)^2}\ , \tag{2}$$

where *N* is the number of travel time data.

4. Cauchy-Steiner Weights and the method of Iteratively Reweighted Least Squares

Robust inversion can be achieved in various ways. To solve the least-absolute-deviation problem, the method of Iteratively Reweighted Least Squares (IRLS) was proposed by Scales [3]. Combining the two methods, the IRLS algorithm can be applied using Cauchy weights, defined as

$$W_{kk} = \frac{\sigma^2}{\sigma^2 + r_k^2}\ , \tag{3}$$

where σ^2 is the scale parameter and r_k is the *k*-th residual. The scale parameter σ^2 of the Cauchy distribution should be known a priori because the data residuals change from iteration to iteration. A condition for the lower bound of the scale parameter was derived by Amundsen [2].

In the framework of Steiner's Most Frequent Value (MFV) method [4] the scale parameter can be determined in an internal iteration. In the *(j+1)*-th step of this procedure the ε_{j+1}^2 (Steiner's scale factor) can be calculated, once ε_j^2 is known, as

$$\varepsilon_{j+1}^2 = 3 \frac{\sum_{k=1}^{N} \frac{r_k^2}{\left(\varepsilon_j^2 + r_k^2\right)^2}}{\sum_{k=1}^{N} \left(\frac{1}{\varepsilon_j^2 + r_k^2} \right)^2} , \tag{4}$$

where the starting value ε_0 in the *0*-th step is given as

$$\varepsilon_0 \leq \frac{\sqrt{3}}{2} \left(r_{max} - r_{min} \right). \tag{5}$$

It can be seen that the above procedure derives the scale parameter from the data set (deviation between measured and calculated data). The stop criterion can be easily defined by experience (for example by a fixed number of iterations). After this the Cauchy-Steiner weights can be calculated by inserting the Steiner's scale parameter (given in the last step of the above internal iterations) into the Cauchy formula (3), which gives the form

$$w_k = \frac{\varepsilon^2}{\varepsilon^2 + r_k^2} . \tag{6}$$

In a Cauchy-Steiner weighted inverse problem the objective function

$$E_w = (\vec{r} , \underline{\underline{W}} \vec{r}) \tag{7}$$

is minimized using the Iteratively Reweighted Least Squares method. In the framework of this algorithm a *0*-th order solution is derived using the (non-weighted) LSQ method and the weights are calculated as

$$w_k^{(0)} = \frac{\varepsilon^2}{\varepsilon^2 + (r_k^{(0)})^2} \tag{8}$$

with

$$r_k^{(0)} = t_k^{measured} - t_k^{(0)} , \tag{9}$$

where the $t_k^{(0)}$ travel times are calculated on the slowness field given by solving the LSQ problem. In the first iteration the objective function

$$E_w^{(1)} = \sum_{k=1}^{N} w_k^{(0)} (r_k^{(1)})^2 \tag{10}$$

is minimized, resulting in the linear set of normal equations

$$\underline{\underline{D}}^T \underline{\underline{W}}^{(0)} \underline{\underline{D}} \, \delta \vec{s} = \underline{\underline{D}}^T \underline{\underline{W}}^{(0)} \delta \vec{t} \tag{11}$$

of the weighted Least Squares method where the $\underline{\underline{W}}^{(0)}$ weighting matrix is of the diagonal form

$$W_{kk}^{(0)} = w_k^{(0)}. \tag{12}$$

Here $\underline{\underline{D}}$ is the distance matrix with the D_{kj} elements giving the length of the ray section in the *j*-th cell belonging to the *k*-th ray, $\vec{s}$ is the slowness vector, $\vec{t}$ is the travel time vector. This procedure is repeated, giving the typical *j*-th iteration step

$$\underline{\underline{D}}^T \underline{\underline{W}}^{(j-1)} \underline{\underline{D}} \, \delta \vec{s} = \underline{\underline{D}}^T \underline{\underline{W}}^{(j-1)} \delta \vec{t} \tag{13}$$

with the $\underline{\underline{W}}^{(j-1)}$ weighting matrix

$$W_{kk}^{(j-1)} = w_k^{(j-1)}. \tag{14}$$

(In these steps the normal equation is linear, because the weights are always calculated in a previous step. Here we note that each step of these iterations contains an internal loop for the determination of the Steiner's scale parameter.) This iteration is repeated until a proper stop criterion is met.

5. CONJUGATE GRADIENT METHOD

In order to solve the normal equations of the type

$$\underline{\underline{D}}^T \underline{\underline{D}} \, \vec{x} = \underline{\underline{D}}^T \vec{b} \,, \tag{15}$$

Scales [1] developed a tomographically very efficient variant of the Conjugate Gradient method. In order to solve the normal equations of the weighted least squares method, this is modified as follows. Let $\vec{x}_o$ be an initial estimate and compute the vectors

$$\vec{s}_o = \vec{b} - \underline{\underline{D}}\,\vec{x}_o, \qquad \vec{r}_o = \vec{p}_o = \underline{\underline{D}}^T \underline{\underline{W}}\left(\vec{b} - \underline{\underline{D}}\,\vec{x}_o\right), \qquad \vec{q}_o = \underline{\underline{D}}\vec{p}_o \tag{16}$$

and start the iteration

$$\alpha_{k+1} = \frac{(\vec{r}_k, \vec{r}_k)}{(\vec{q}_k, \vec{q}_k)} \tag{17}$$

$$\vec{x}_{k+1} = x_k + \alpha_{k+1}\vec{p}_k \tag{18}$$

$$\vec{s}_{k+1} = \vec{s}_k - \alpha_{k+1}\vec{q}_k \tag{19}$$

$$\vec{r}_{k+1} = \underline{\underline{D}}^T \underline{\underline{W}}\,\vec{s}_{k+1} \tag{20}$$

$$\beta_{k+1} = \frac{(\vec{r}_{k+1}, \vec{r}_{k+1})}{(\vec{r}_k, \vec{r}_k)} \tag{21}$$

$$\vec{p}_{k+1} = \vec{r}_{k+1} + \beta_{k+1}\vec{p}_k \tag{22}$$

$$\vec{q}_{k+1} = \underline{\underline{D}}\vec{p}_{k+1} \tag{23}$$

where k = 0,1,2... refers to the iteration number. This procedure differs from the ordinary Conjugate Gradient algorithm only at two points, where beside the transpose of the $\underline{\underline{D}}$ matrix, also the $\underline{\underline{W}}$ weight matrix appears.

5.1. Conjugate Gradients with Gaussian dataset

At first the original (non-weighted) CG algorithm was applied for reconstruction of Dataset I. The resulting velocity distribution is shown in *Figure 2*.

The model distance is around 6%, which is quite an acceptable result. The image shows a relatively good reconstruction, as the CG algorithm minimizes the L_2 norm and solves the normal equation of the LSQ method, although the dataset contains 1% Gaussian noise.

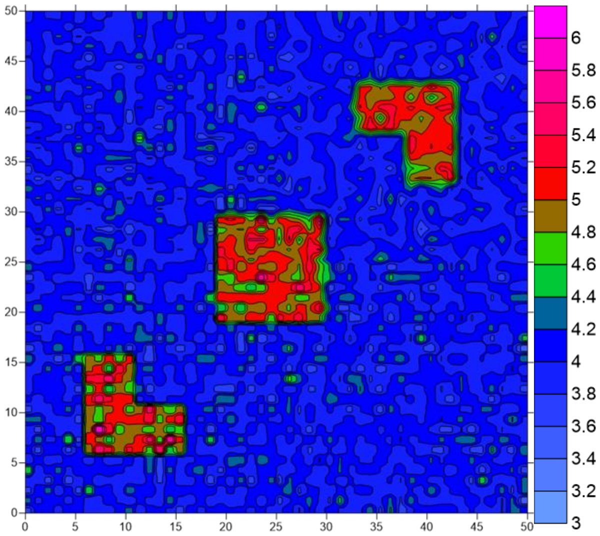

Figure 2. *Tomographic CGRAD inversion of Dataset I (data distance: 0.00948; model distance: 0.0579)*

5.2. Weighted Conjugate Gradients (W-CGRAD) with Gaussian data

If the ordinary Weighted Conjugate Gradients (W-CGRAD) algorithm is applied to Gaussian data (Dataset I), the velocity distribution looks as presented in *Figure 3*.

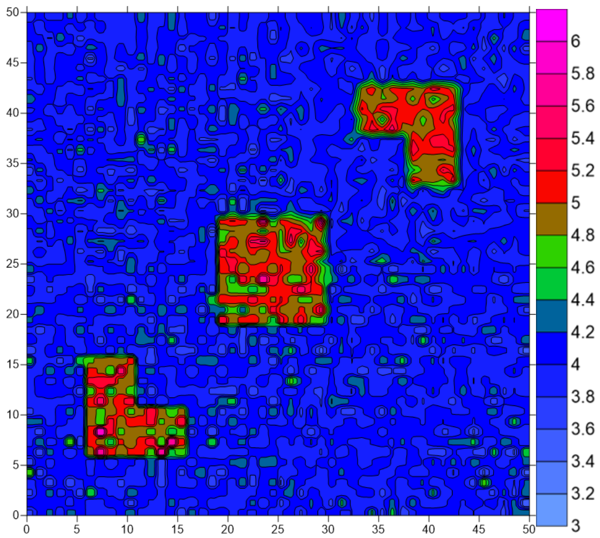

Figure 3. *Tomographic W-CG inversion of Dataset 1 (model distance: 0.0641)*

The Conjugate Gradients result in *Figure 2* is slightly better, because the W-CGRAD method solves the Gaussian least squares problem and the dataset follows a Gaussian distribution (the optimal method is the LSQ without weighting).

5.3. Conjugate Gradients with non-Gaussian dataset

If the ordinary Conjugate Gradients (CGRAD) algorithm is applied to non-Gaussian data (Dataset II), the velocity distribution looks as presented in *Figure 4*.

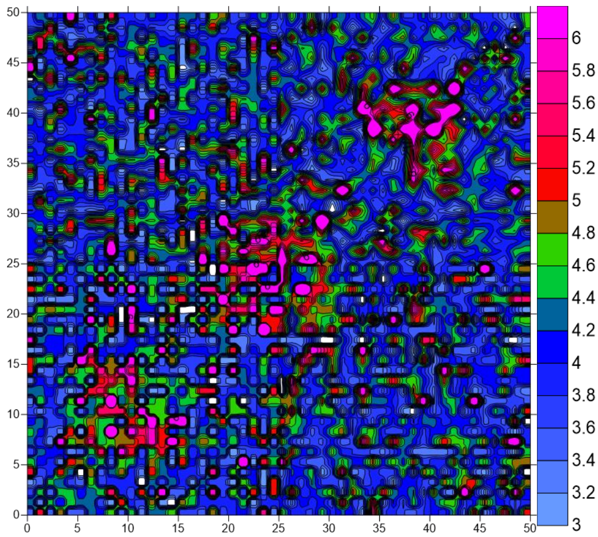

Figure 4. *Ordinary tomographic CGRAD inversion of Dataset II containing outliers (model distance: 0.250)*

In this case the model distance is 25%, which is really high for the tomography model. The image is highly distorted and almost unrecognizable. This result shows that the simple Gaussian Least Square Method is very sensitive to non-Gaussian noises, for example outliers in the datasets.

5.4. Weighted Conjugate Gradients with non-Gaussian dataset

If the ordinary Conjugate Gradients algorithm is modified to solve the normal equations of the Weighted LSQ method with the use of Cauchy-Steiner Weights, a robust tomography method can be defined (W-CGRAD). *Figure 5* shows the reconstructed image where the weighted version of CG algorithm is applied to process Dataset II. It can be seen that the influence of outliers is highly reduced due to the weighting and a highly improved image is obtained. This result shows that the use of Cauchy-Steiner weights is very efficient in reducing the influence of outliers.

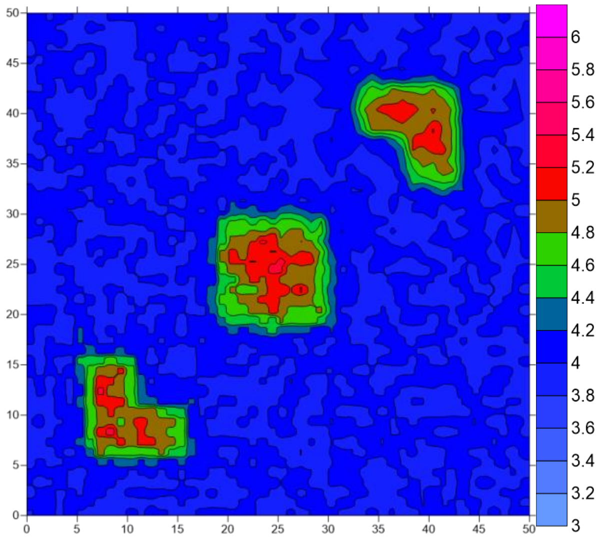

Figure 5. *The tomographic reconstruction of Dataset II (containing outliers) using the weighted W-CGRAD method (Model distance: 0.0871)*

6. Simultaneous Iterative Reconstruction Technique (SIRT)

Firstly the back projection method was the first to be used for seismic tomography. Later the Least Square problems in tomography came to be solved by Algebraic Reconstruction Technique (ART) and Simultaneous Iterative Reconstruction Technique (SIRT). SIRT remains one of the most frequently used iterative methods in seismic tomography with the improvement formula

$$x_j^{(q+1)} = x_j^{(q)} + \frac{1}{Q_j} \sum_{i=1}^{Q_j} \frac{D_{ij} r_i^{(q)}}{\sum_k D_{ik}^2}. \qquad (24)$$

Q_j is the number of traces running through the *j*-th cell, $r_i^{(q)}$ means the difference between the *i*-th measured and calculated data and D_{ij} is the ray section of the *i*-th ray in the *j*-th cell.

As can be seen, the SIRT formula calculates the arithmetic mean of the ART corrections calculated for the given (*j*-th) cell. It is well known that arithmetic mean is efficient only if the data noise follows Gaussian distribution. In our case, we also have non-Gaussian noise distribution. Therefore, we have to use the weighted mean instead of arithmetic one

$$x_j^{(q+1)} = x_j^{(q)} + \frac{1}{\sum_{l=1}^{Q_j} W_{ll}} \sum_{i=1}^{Q_j} W_{ii} \frac{D_{ij} r_i^{(q)}}{\sum_k D_{ik}^2}, \tag{25}$$

where Q_j is the number of traces covered in the *j*-th cell, $r_i^{(q)}$ means the difference between measured and calculated data, D_{ij} is the ray section of the *i*-th ray trace in the *j*-th cell and W_{ii} is the Cauchy-Steiner weight belonging to the *i*-th travel time.

6.1. Simultaneous Iterative Reconstruction Technique (SIRT) with Gaussian dataset

It is well known that the SIRT method is one of the best methods in tomography when the distribution of noises follows Gaussian statistics. This can be proved by using Dataset I in a SIRT reconstruction.

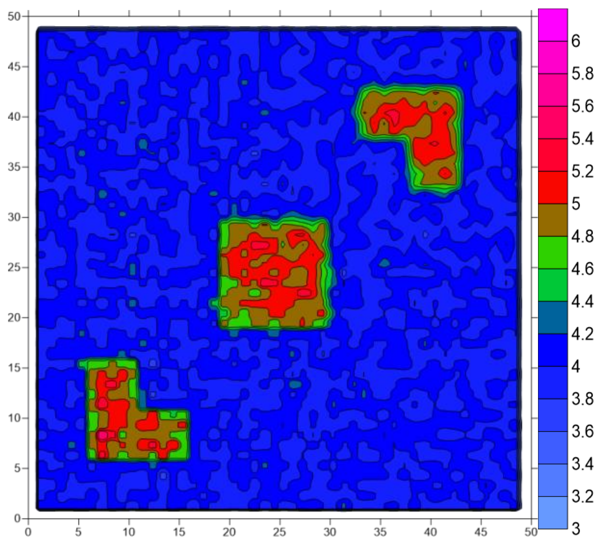

Figure 6. *Tomographic SIRT inversion of tDataset I (data distance: 0.00973; model distance: 0.0216)*

Figure 6 shows that the SIRT method gives nearly three times better model distance compared to the results of CG method.

6.2. Weighted Simultaneous Iterative Reconstruction Technique (W-SIRT) with Gaussian dataset

As shown in *Figure 7*, using Dataset I the W-SIRT method gives similar reconstruction to that given by the traditional SIRT *(Figure 6)*, which seems to be slightly better; however, there is only a negligible difference in the model distances.

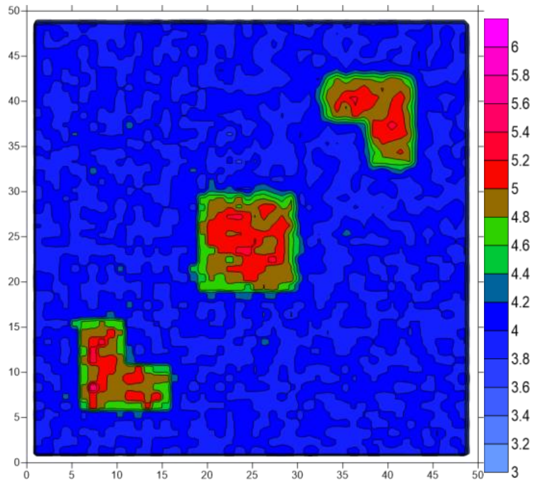

Figure 7. *Tomographic W-SIRT inversion of Dataset I (model distance: 0.0227)*

6.3. Simultaneous Iterative Reconstruction Technique with non-Gaussian dataset

The SIRT method produces a relatively distorted image in the case of Dataset II, containing outliers. This is demonstrated in *Figure 8*.

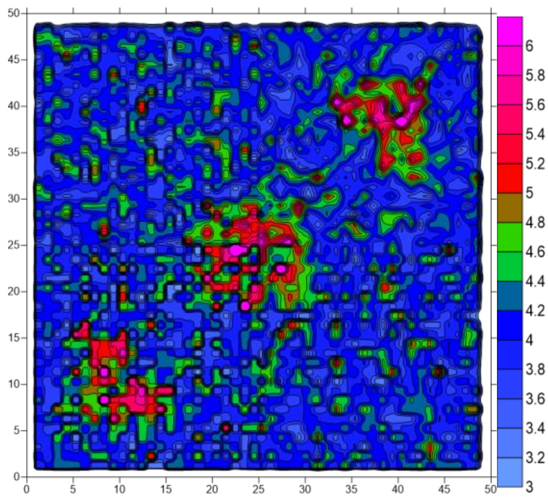

Figure 8. *The tomographic SIRT inversion of Dataset II (model distance: 0.0635)*

As can be seen, the SIRT method is also very sensitive to the non-Gaussian nature of the noise, though it is better than CG with the non-Gaussian dataset (the model distance for CG with non-Gaussian data was 0.250). In spite of this fact, the reconstruction in *Figure 8* is not acceptable.

6.4. Weighted Simultaneous Iterative Reconstruction Technique (W-SIRT) with non-Gaussian dataset

If we use the W-SIRT method for reconstructing Dataset II, we obtain the velocity distribution shown in *Figure 9*. It can be seen that the algorithm using Cauchy-Steiner weights is highly resistant to outlier data and gives a good reconstruction result.

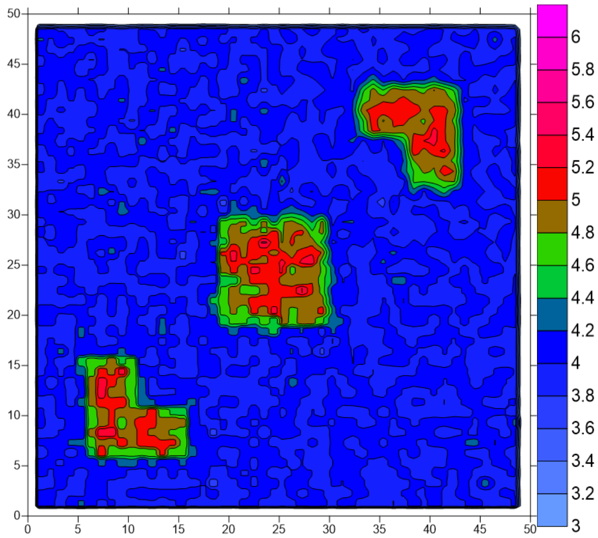

Figure 9. *The tomographic reconstruction of Dataset II by the W-SIRT method (model distance: 0.0242)*

The model distance is equal to 0.0242, which is nearly the same as when the SIRT method was applied on the Gaussian dataset, where the model distance was 0.026. Of course, SIRT with the Gaussian dataset should give a better result than W-SIRT with the non-Gaussian dataset. Investigations with outlier data proved that the W-SIRT algorithm is computationally noise resistant and computationally economic.

7. CONCLUSIONS

The proposed tomography algorithms are tested for a medium-size tomography example using synthetic travel time data. It is proved that, compared to their traditional versions, the outlier sensitivity of the generalized tomography methods is sufficiently reduced. Compared to the traditional CG algorithm the new W-CGRAD algorithm is more robust and resistant to outliers. The SIRT method was also modified by using

Cauchy-Steiner weights and it was proved that, compared to its original version, the W-SIRT method is less sensitive to outlier data.

ACKNOWLEDGEMENTS

The research work of author Utku Kale was supported by the Stipendium Hungaricum Scholarship Program. The research of Mihály Dobróka was supported by the National Research, Development and Innovation Office – NKFIH, Project No. K 109441.

REFERENCES

[1] SCALES, J. A.: Tomographic inversion via the conjugate gradient method. *Geophysics*, 1987, 52, 179–185.

[2] AMUNDSEN, L.: Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion. *Geophysics*, 1991, 56:(12), 2027–2038.

[3] SCALES, J. A.–GERSZTENKORN, A.–TREITEL, S.: Fast l_p solution of large, sparse, linear systems: Application to seismic travel time tomography. *Journal of Computational Physics*, 1988, 75. 314–333.

[4] STEINER, F.: Most frequent value procedures. (A short monograph). *Geophysical Transactions*, 1988, 34 (2–3), 139–260.