

On the Generalization of Seismic Tomography Algorithms

Mihály Dobróka^{1,2}, Hajnalka Szegedi²

¹MTA-ME Applied Geosciences Research Group, Miskolc, Hungary

²Department of Geophysics, University of Miskolc, Miskolc, Hungary

Email: dobroka@uni-miskolc.hu

Received November 13, 2013; revised December 13, 2013; accepted December 18, 2013

Copyright © 2014 Mihály Dobróka, Hajnalka Szegedi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Mihály Dobróka, Hajnalka Szegedi. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

ABSTRACT

The seismic tomography problem often leads to underdetermined and inconsistent system of equations. Solving these problems, care must be taken to keep the propagation of data errors under control. Especially, the non-Gaussian nature of the noise distribution (for example outliers in the data sets) can cause appreciable distortions in the tomographic imaging. In order to reduce the sensitivity to outlier, some generalized tomography algorithms are proposed in the paper. The weighted version of the Conjugate Gradient method is combined with the Iteratively Reweighted Least Squares (IRLS) procedure leading to a robust tomography method (W-CGRAD). The generalized version of the SIRT method is introduced in which the (Cauchy-Steiner) weighted average of the ART corrections is used. The proposed tomography algorithms are tested for a small sized tomography example by using synthetic traveltimes data. It is proved that—compared to their traditional versions—the outlier sensitivities of the generalized tomography methods are sufficiently reduced.

KEYWORDS

Robust Tomography; Conjugated Gradients; SIRT; Cauchy-Steiner Weights

1. Introduction

The classic geophysical tomography problem is solved to reconstruct the velocity distribution for the investigated portion of the Earth such that the projected data (the traveltimes) should agree with measurements. In practice, this is done by solving a large, sparse, least-squares problem. 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 proved by [1] that the Conjugate Gradient (CG) method can be used even in large-scale tomographic least squares inversion with taking the advantage of the sparsity of the matrix of the problem. This low cost CG algorithm of Scales was applied in solving the double trace (DT) tomography problem [2].

It is well-known that least squares methods (LSQ) give optimal results only when the data noises follow a Gaussian distribution, or in other words the least squares solution is very sensitive to sparsely distributed large errors, i.e. outliers in the data set. The practice shows that the distribution of the errors in the measured data is seldom Gaussian so that the use of the least squares method cannot be optimal.

There are various ways to address the question of the statistical robustness. One of the most frequently used methods in robust optimization is the Least Absolute Deviation (LAD) method. In this case, L1 norm is used to characterize the misfit between the observed and predicted data. The inversion with minimization of the L1 norm gives more reliable estimates when a small number of large errors contaminate the data. Another possibility is to use the Cauchy criterion. In this case, the Cauchy distribution of data noise is assumed. Cauchy inversion is also frequently used in the geophysical inversion as a robust optimization method [3]. However, there is a problem with Cauchy inversion: the scale parameter of the weights should be a priori known. An improved algorithm was developed by Steiner (1988) in the framework of the method of Most Frequent Values (MFV) [4]

in which the scale parameters are automatically derived from the data set.

As it is well-known, various tools of image processing are applicable in seismic tomography. It was proved by [5] that smoothing the tomograms (at the end of the reconstruction) by the use of the so-called alpha-trimmed mean, the distortions caused by the outliers can appreciably be reduced. In this paper we report on generalization of two items, in that we apply Cauchy-Steiner weights (Cauchy weights with scale parameters automatically determined by using Steiner's MFV method) in the original Conjugate Gradient tomography algorithm in order to minimize the weighted norm of the deviation vector. The generalization of the Simultaneous Iterative Reconstruction Technique (SIRT) is proposed and in the data space Cauchy-Steiner weights are used to calculate the weighted average of the ART corrections (W-SIRT).

2. The IRLS Technique Using Cauchy-Steiner Weights

The realization of robust inversion can be achieved in various ways. To handle the Cauchy inversion problem the Gradient Method was used by [3]. To solve the least-absolute-deviation problem, the method of Iteratively Reweighted Least Squares (IRLS) was proposed by [6]. Combining the two methods the IRLS algorithm can be applied using Cauchy weights, defined as

$$W_{kk} = \frac{\sigma^2}{\sigma^2 + r_k^2}, \quad (1)$$

where σ^2 is the scale parameter and r_k is k -th residual. The scale parameter σ^2 of the Cauchy distribution cannot be given a priori because the data residuals change from iteration to iteration. A condition for the lower bound of the scale parameter was derived by [3].

In the framework of Steiner's (MFV) method [4] the scale parameter σ^2 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 in the knowledge of ε_j^2 as

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

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

$$\varepsilon_0 \leq \frac{\sqrt{3}}{2} (r_{\max} - r_{\min}). \quad (3)$$

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 experiences (for example by 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 (1) which gives the form

$$w_k = \frac{\varepsilon^2}{\varepsilon^2 + r_k^2}. \quad (4)$$

The same weights were used in [7] for the joint inversion of in-mine measured geoelectric and seismic data and recently in [8] for creating robust inversion based Fourier transformation.

In order to minimize the weighted norm

$$E_w = (\bar{r}, \underline{\underline{W}}\bar{r}) \quad (5)$$

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

$$w_k^{(0)} = \frac{\varepsilon^2}{\varepsilon^2 + (r_k^{(0)})^2} \quad (6)$$

with

$$r_k^{(0)} = t_k^{\text{measured}} - t_k^{(0)}, \tag{7}$$

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

$$E_w^{(1)} = \sum_{k=1}^N w_k^{(0)} \left(r_k^{(1)} \right)^2 \tag{8}$$

is minimized resulting in the linear set of normal equations

$$\underline{\underline{D}}^T \underline{\underline{W}}^{(0)} \underline{\underline{D}} \delta \vec{s}^{(1)} = \underline{\underline{D}}^T \underline{\underline{W}}^{(0)} \delta \vec{t}. \tag{9}$$

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

$$W_{kk}^{(0)} = w_k^{(0)}. \tag{10}$$

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, \vec{t} is the traveltime. This procedure is repeated giving the typical j -th iteration step

$$\underline{\underline{D}}^T \underline{\underline{W}}^{(j-1)} \underline{\underline{D}} \delta \vec{s} = \underline{\underline{D}}^T \underline{\underline{W}}^{(j-1)} \delta \vec{t} \tag{11}$$

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

$$W_{kk}^{(j-1)} = w_k^{(j-1)}. \tag{12}$$

(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 contain an internal loop for the determination of the Steiner's scale parameter.) This iteration is repeated until a proper stop criterion is met.

In order to solve the normal equations of the type

$$\underline{\underline{D}}^T \underline{\underline{D}} \vec{x} = \underline{\underline{D}}^T \vec{b}, \tag{13}$$

[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, it should be modified as follows:

Let \vec{x}_o be an initial estimate of the slowness and compute the vectors

$$\vec{s}_o = \vec{b} - \underline{\underline{D}} \vec{x}_o, \quad \vec{r}_o = \vec{p}_o = \underline{\underline{D}}^T \underline{\underline{W}} \left(\vec{b} - \underline{\underline{D}} \vec{x}_o \right), \tag{14}$$

$$\vec{q}_o = \underline{\underline{D}} \vec{p}_o \tag{15}$$

and start the iteration

$$\alpha_{k+1} = \frac{(\vec{r}_k, \vec{r}_k)}{(\vec{q}_k, \vec{q}_k)} \tag{16}$$

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_{k+1} \vec{p}_k \tag{17}$$

$$\vec{s}_{k+1} = \vec{s}_k - \alpha_{k+1} \vec{q}_k \tag{18}$$

$$\vec{r}_{k+1} = \underline{\underline{D}}^T \underline{\underline{W}} \vec{s}_{k+1} \tag{19}$$

$$\beta_{k+1} = \frac{(\vec{r}_{k+1}, \vec{r}_{k+1})}{(\vec{r}_k, \vec{r}_k)} \tag{20}$$

$$\vec{p}_{k+1} = \vec{r}_{k+1} + \beta_{k+1} \vec{p}_k \tag{21}$$

$$\vec{q}_{k+1} = \underline{\underline{D}} \vec{p}_{k+1} \tag{22}$$

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

For the numerical experiments a rectangular test area of the size 100×100 cells was defined (**Figure 1**). In the model three anomalies of the velocity 5 km/s (red color) is located in a homogeneous background of 4 km/s velocity (blue color). Sources and receivers were positioned along the x and y axis in an arrangement fulfilling the requirement of full tomographic ray coverage, so the theoretical traveltimes data were computed along 60,000 ray traces.

Two data sets were generated for the tests. In the data set I. random noises up to 1% of theoretical traveltimes—following the Gauss distribution—were added to the data. The second data set—containing outliers—(data set II.) was created from the first one so that 20% extra noise was added to a randomly selected 20% portion of the data. In our numerical tests 10 iterations were applied in CG reconstructions.

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

$$D_{mod} = \sqrt{\frac{1}{M} \sum_{j=1}^M \left(\frac{x_j - x_j^{(0)}}{x_j^{(0)}} \right)^2} \quad (23)$$

was used. Here x_j and $x_j^{(0)}$ denotes the slowness in the j -th cell of the reconstructed picture and the model, respectively, M is the number of cells. Similarly the relative data distance were also calculated

$$D_{dat} = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(\frac{t_k^{(input)} - t_k^{(reconstructed)}}{t_k^{(input)}} \right)^2}. \quad (24)$$

Using the first data set, the ordinary CG algorithm was applied for reconstruction results the velocity distribution shown on **Figure 2**. We got reliable results with acceptable relative distances both in data space and in model space.

The weighted version of the conjugated gradients (W-CGRAD) gives similar result, as it is shown in **Figure 3**. The CG result in **Figure 2** is slightly better, because this method solves the Gaussian least squares problem and the data set contains Gaussian noise.

In order to test the outlier-sensitivity of the ordinary (non-weighted) CG method let us now use non-Gaussian data (data set II.). The reconstructed velocity distribution is given by **Figure 4**. As it can be seen, the picture is highly distorted and very noisy demonstrating the fact, that the ordinary Conjugate Gradient method (giving a least squares solution) is very sensitive for the outlier. This is an obvious consequence of the fact that the ordinary CGRAD algorithm of [1] is a variant of the Least Squares method which gives optimal result only in the case of Gaussian data noises.

Note, that the relative distances are very large, which is in agreement with the fact, that the anomalies are almost unrealizable.

Let us now test the newly proposed weighted conjugate gradient algorithm (W-CGRAD) using the Cauchy-Steiner weights. **Figure 5** shows the reconstructed picture where the weighted version of the CG algorithm was applied to process data set II. (containing the outliers). It can be seen, that the influence of outliers is highly reduced due to the weighting. Note, that the relative model distance (0.103) is approximately seven times smaller, than in the case of **Figure 4**, where the original (non-weighted) CG algorithm was used. This result shows that the weighting with Cauchy-Steiner weights is very efficient in reducing the influence of outliers.

3. A Cauchy-Steiner Weighted SIRT Method

Simultaneous Iterative Reconstruction Technique is one of the most frequently used methods in seismic tomography. In the typical step of the algorithm the arithmetic mean of the so-called ART correction belonging to the Q seismic rays crossing the j -th cell is calculated as

$$x_j^{(q+1)} = x_j^{(q)} + \frac{1}{Q_j} \sum_{i=1}^{Q_j} \frac{D_{ij} r_i^{(q)}}{\sum_k D_{ik}^2}. \quad (25)$$

If instead of this simple average, a weighted average is used

$$x_j^{(q+1)} = x_j^{(q)} + \frac{1}{\sum_{l=1}^{Q_j} W_{ll}} \sum_{i=1}^{Q_j} W_{ii} \frac{D_{ij} r_i^{(q)}}{\sum_k D_{ik}^2} \quad (26)$$

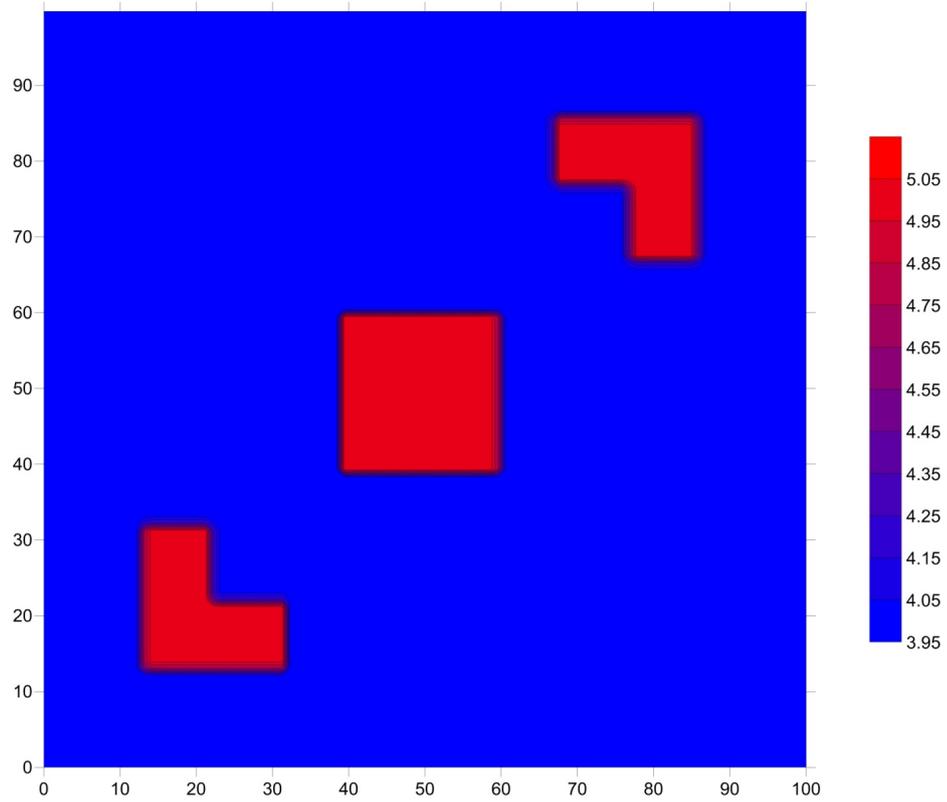


Figure 1. The model used for numerical tests.

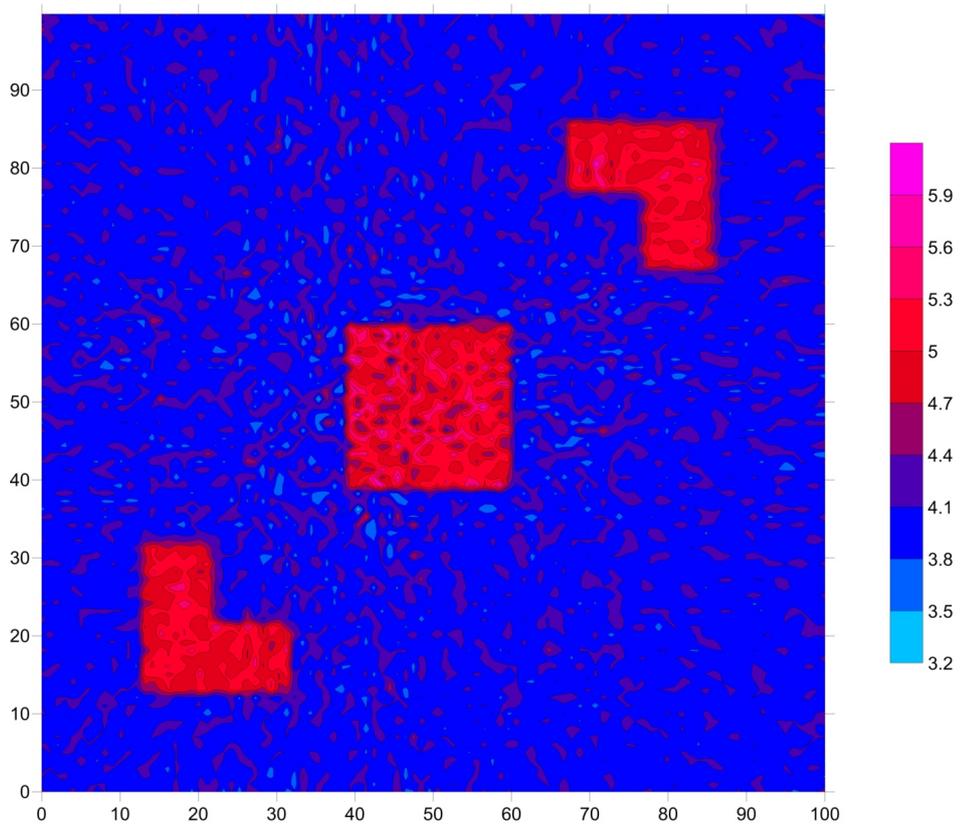


Figure 2. Tomographic CGRAD inversion of the data set I. (Data distance: 0.00948, Model distance: 0.0579)

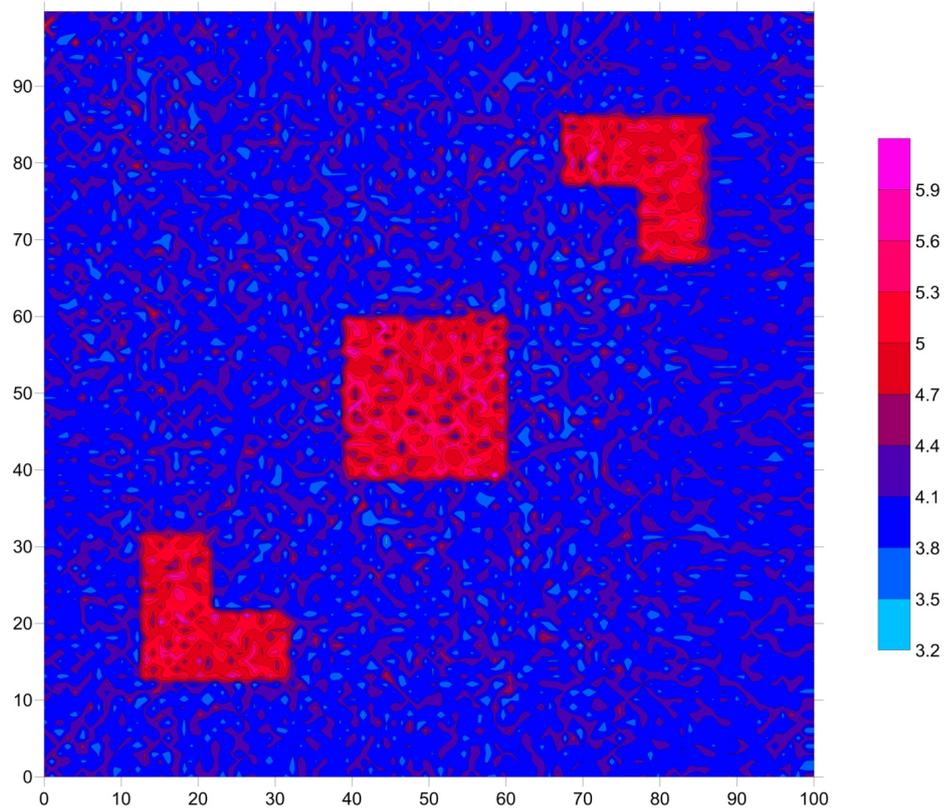


Figure 3. Tomographic W-CGRAD inversion of the data set I. (Model distance: 0.0641)

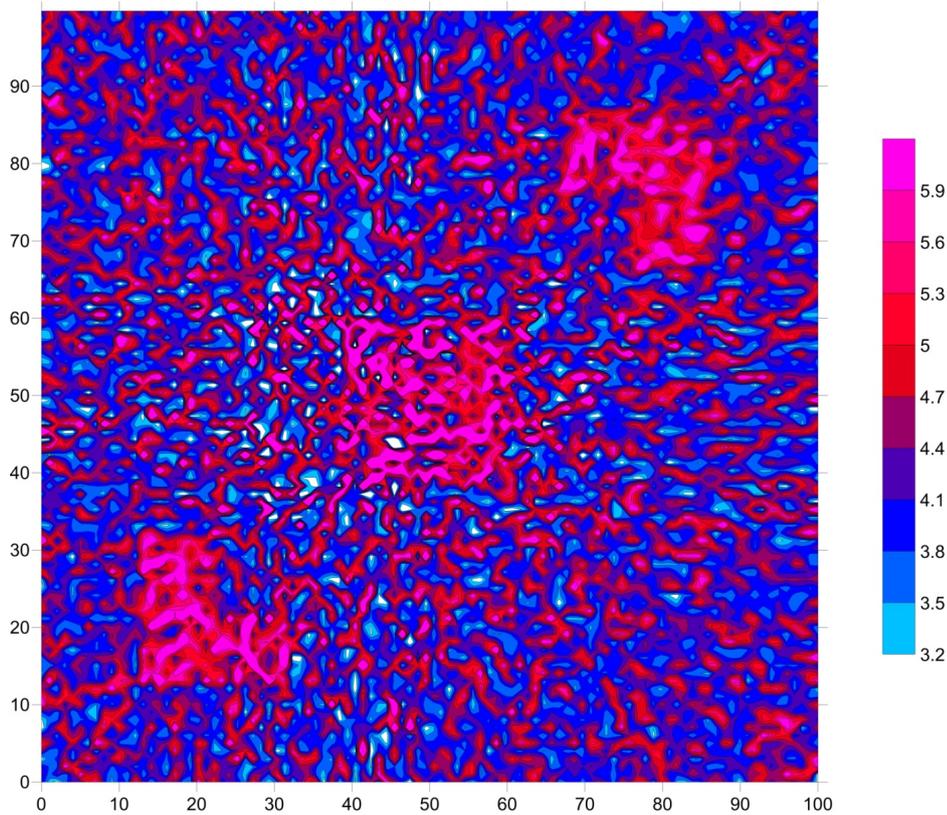


Figure 4. Ordinary tomographic CGRAD inversion of the data set II. containing outliers. (Model distance: 0.250)

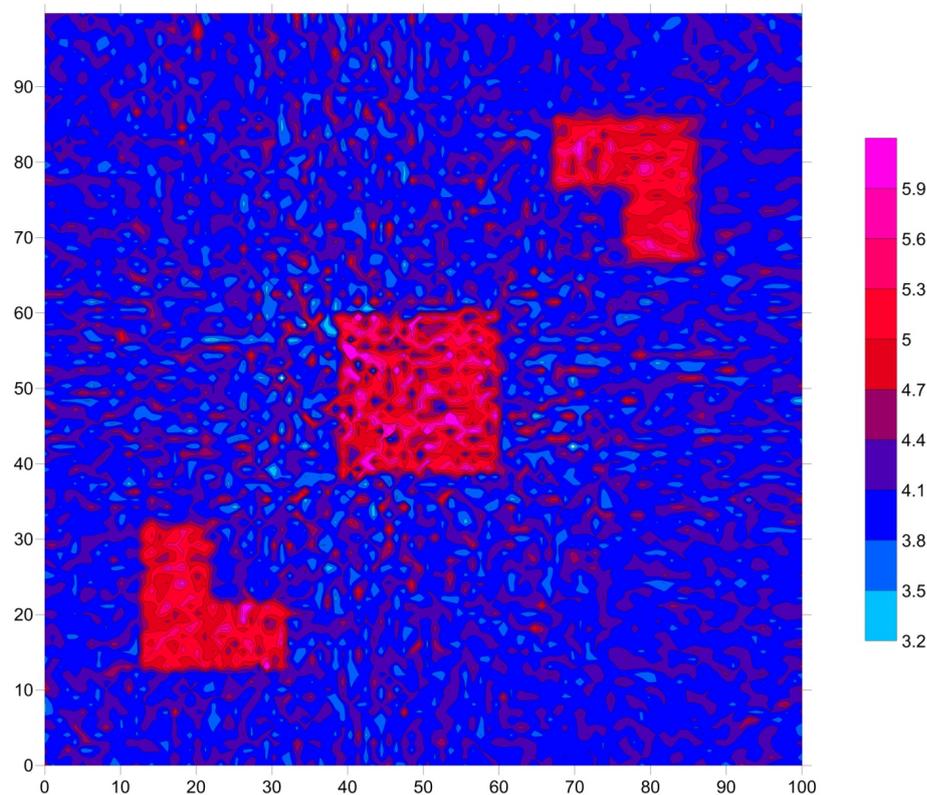


Figure 5. The tomographic reconstruction of the data set II. (containing outliers) using the weighted W-CGRAD method. (Model distance: 0.0871)

a new version of the SIRT algorithm can be defined. Using Cauchy-Steiner weights a robust SIRT method can be expected.

As it is well-known, the SIRT method is one of the best procedures in tomography when the noise distribution follows Gaussian statistics. This can also be proved by using data set I. in a SIRT reconstruction. The result is shown in **Figure 6**. It can be seen that the compared to CG, the SIRT method gives nearly two times better model distance.

The weighted version of the SIRT (W-SIRT) gives similar result, as it is shown in **Figure 7**. The SIRT result in **Figure 6** is slightly better, but there is only negligible difference in the model distance.

Similarly to the CG procedure the traditional SIRT method produce relatively distorted picture in case of data sets containing outliers. This is demonstrated in **Figure 8**, where the SIRT reconstruction by using data set II. is shown. Though this picture is better than the CG result given in **Figure 4** (the model distance is reduced by a factor of two), the reconstruction is not acceptable.

On the other hand, using the weighted version of the SIRT method, we obtained the velocity distribution shown by **Figure 9**. It can be seen, that the weighted SIRT algorithm with Cauchy-Steiner weights (W-SIRT) is highly resistant to outlier data. Comparing the model distances of **Figure 6** and **Figure 8** we can see that the introduction of Cauchy-Steiner weights results in sufficient rejection of the influence of outliers. Note it that the modification required by the weighting can easily be implemented into an ordinary SIRT algorithm and—due to the weighting—the computation time is increased by only a very small amount. So, the W-SIRT algorithm is computationally economic and noise resistant.

4. Conclusion

In order to make the Conjugate Gradient and the SIRT tomography methods more robust a new weight factor, the Cauchy-Steiner weights were introduced by calculating the scale parameter of the Cauchy weight using Steiner's method of the Most Frequent Values. Minimizing the (Cauchy-Steiner) weighted norm of deviations the Conjugate Gradient algorithm (W-CGRAD) was applied in the framework of the Iteratively Reweighted Least Squares method. Compared to the Gaussian Least Squares method, it was demonstrated in small sized to

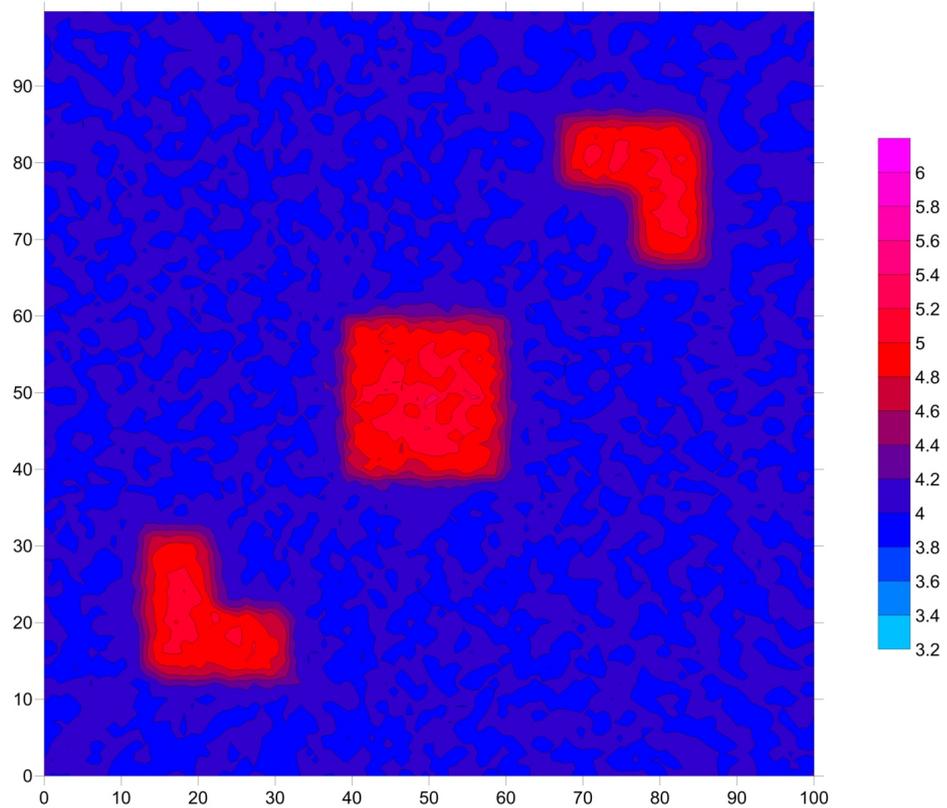


Figure 6. Tomographic SIRT inversion of the data set I. (Data distance: 0.00973, Model distance: 0.0216)

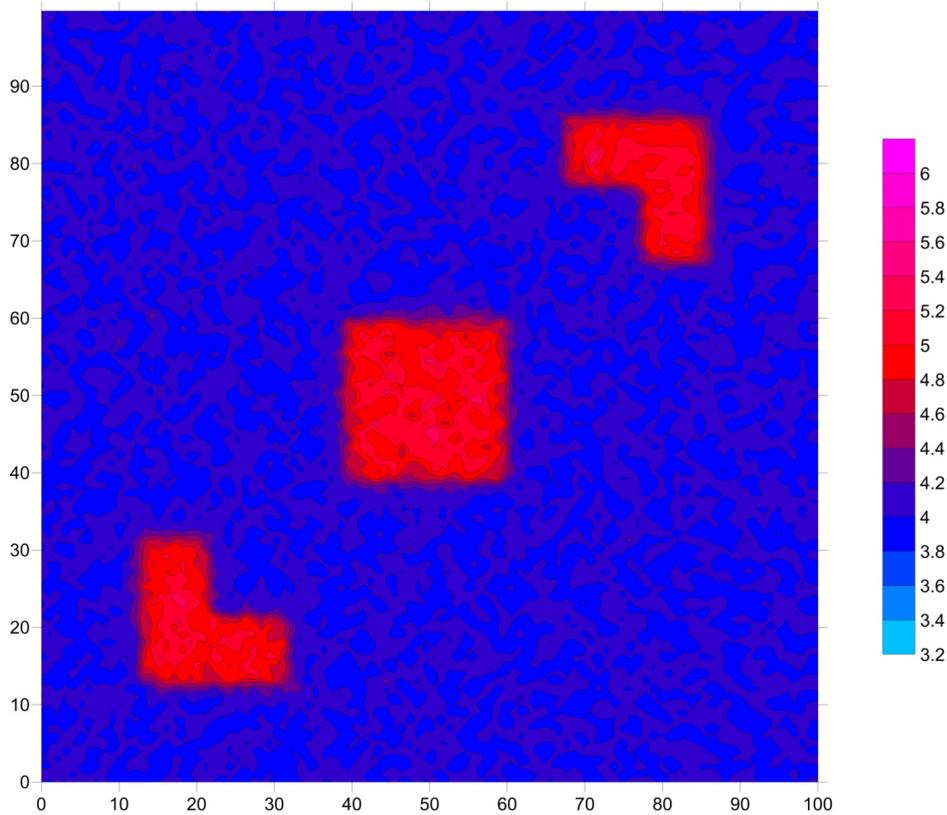


Figure 7. Tomographic W-SIRT inversion of the data set I. (Model distance: 0.0227)

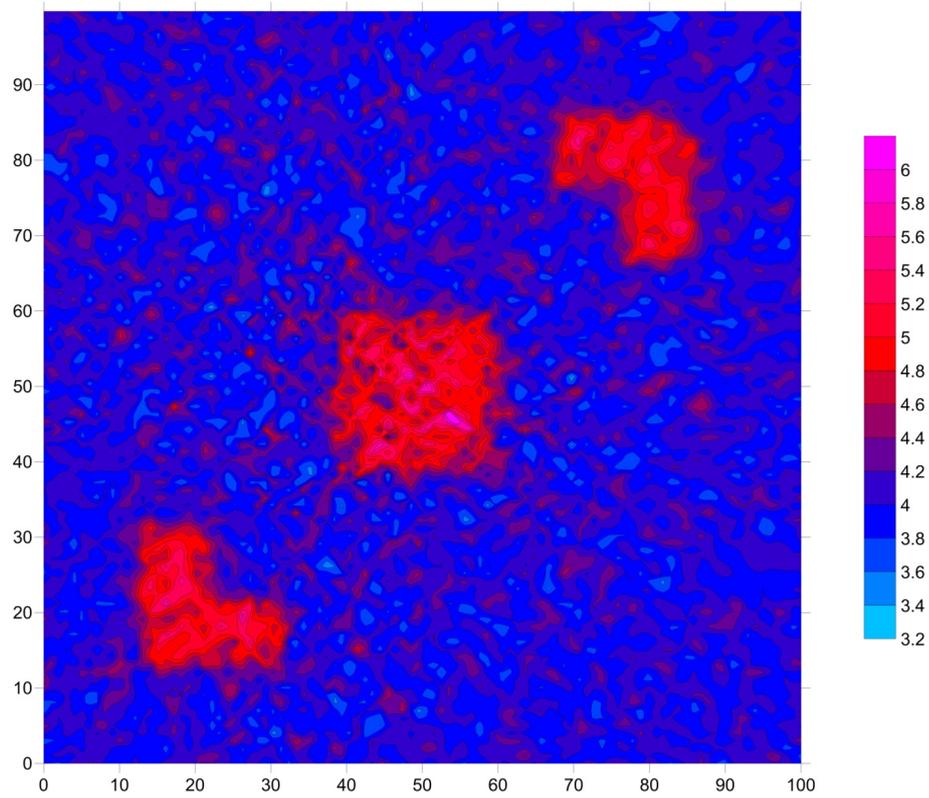


Figure 8. Tomographic SIRT inversion of the data set II. (Model distance: 0.0635)

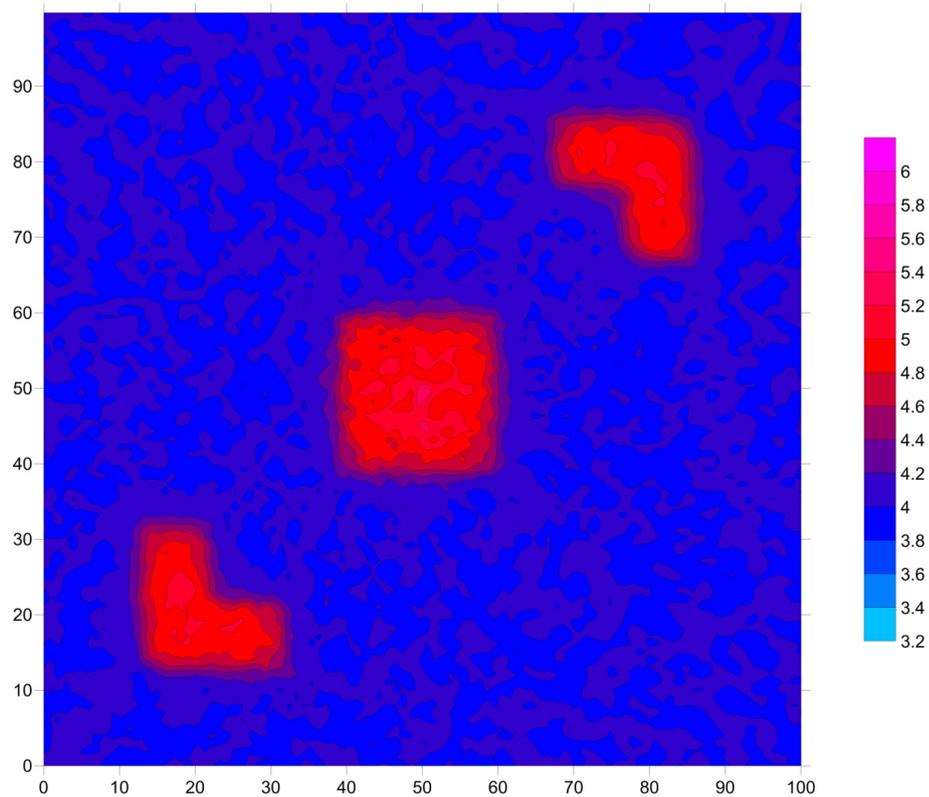


Figure 9. The tomographic reconstruction of the data set II. (containing outliers) using the weighted W-SIRT method. (Model distance: 0.0242)

mographic experiment, that the new (W-CGRAD) algorithm is more robust and resistant for outliers. The SIRT method was also modified by using Cauchy-Steiner weights. Instead of arithmetic mean of the ART corrections, the W-SIRT algorithm uses weighted average in calculating the new slowness field. It was also proved that—compared to its original version—the W-SIRT method is less sensitive for outlier data. Using the robust tomography methods proposed in the paper, reduced outlier sensitivity can be achieved resulting in better tomographic reconstruction.

Acknowledgements

The investigations were carried out in the framework of research program of the MTA-ME Applied Geosciences Research Group and were partly supported by the OTKA (project No. K109441). The second author was supported from the TÁMOP 4.2.4. A/2-11-1-2012-0001, National Excellence Program—Elaborating and operating an inland student and researcher personal support system convergence program. The project was subsidized by the European Union and co-financed by the European Social Fund. The authors thank the support.

REFERENCES

- [1] J. A. Scales, “Tomographic Inversion via the Conjugate Gradient Method,” *Geophysics*, Vol. 52, No. 2, 1987, pp. 179-185. <http://dx.doi.org/10.1190/1.1442293>
- [2] M. Dobroka, L. Dresen, C. Gelbke and H. Rüter, “Tomographic Inversion of Normalized Data: Double-Trace Tomography Algorithms,” *Geophysical Prospecting*, Vol. 40, No. 1, 1992, pp. 1-14. <http://dx.doi.org/10.1111/j.1365-2478.1992.tb00361.x>
- [3] L. Amundsen, “Comparison of the Least-Squares Criterion and the Cauchy Criterion in Frequency-Wave Number Inversion,” *Geophysics*, Vol. 56, No. 12, 1991, pp. 2027-2038. <http://dx.doi.org/10.1190/1.1443015>
- [4] F. Steiner, “Most Frequent Value Procedures,” *Geophysical Transactions*, Vol. 34, No. 2-3, 1988, pp. 139-260.
- [5] A. Gersztenkorn and J. A. Scales, “Smoothing Seismic Tomograms with Alpha-Trimmed Means,” *Geophysical Journal*, Vol. 92, No. 1, 1988, pp. 67-72. <http://dx.doi.org/10.1111/j.1365-246X.1988.tb01121.x>
- [6] J. A. Scales, A. Gersztenkorn and S. Treitel, “Fast Lp Solution of Large, Sparse, Linear Systems: Application to Seismic Travel Time Tomography,” *Journal of Computational Physics*, Vol. 75, No. 2, 1988, pp. 314-333. [http://dx.doi.org/10.1016/0021-9991\(88\)90115-5](http://dx.doi.org/10.1016/0021-9991(88)90115-5)
- [7] M. Dobroka, A. Gyulai, T. Ormos, J. Csókás and L. Dresen, “Joint Inversion of Seismic and Geoelectric Data Recorded in an Under-Ground Coal Mine,” *Geophysical Prospecting*, Vol. 39, No. 5, 1991, pp. 643-665. <http://dx.doi.org/10.1111/j.1365-2478.1991.tb00334.x>
- [8] H. Szegedi and M. Dobroka, “On the Use of Steiner’s Weights in Inversion-Based Fourier Transformation,” *Acta Geodaetica et Geophysica*, 2014, pp. 1-10. <http://dx.doi.org/10.1007/s40328-014-0041-0>