

Algebraic Reconstruction Technique for Neurofuzzy Geotomography

Tetsuya Miyoshi, Hajime Tabuchi and Hidetomo Ichihashi

Department of Industrial Engineering

Osaka Prefecture University

1-1 Gakuen-cho, Sakai, Osaka 593, JAPAN

miyoshi@ie.osakafu-u.ac.jp

Abstract

A new Algebraic Reconstruction Techniques(ART) are developed for a neurofuzzy geotomography to accelerate the convergence of learning phase and to reduce the learning time or iteration times. The learning algorithm is derived from a constrained optimization problem. Minkowski norm of the corrections of parameters is used as the objective function of the optimization problem. Some computer simulation results show that smooth distributions of a material parameter are obtained by using the Minkowski norm. Furthermore, the proposed method is applied to the experimental data collected at a dam site by cross borehole seismic probing.

1. Introduction

Computerized Tomography(CT) is one of the methods to visualize a spatial distribution of material parameters, such as attenuation rate or propagation velocity, from measured projection data. Geotomography(Geophysical Tomography) has been used in geological and mineral prospecting applications and can image geological structures using some iterative techniques. Algebraic Reconstruction Techniques(ART) were introduced by Gordon *et al.*[4] for solving the problem of three dimensional reconstruction from projections in electron microscopy and radiology. Variations on the ART[5, 6, 7, 11] have been proposed and all ART algorithms share a property of iteratively attempting to match the weighted or unweighted sums of appropriate reconstruction elements with the corresponding projection data. Dines and Lytle proposed the iterative method for geophysical tomography by using the ART and analyzed synthetic and experimental data [3].

The ART and the normalized least mean square (NLMS) are the similar algorithms of iterative learning. The NLMS also is a method to accelerate the convergence of the learning phase[1]. It is only the learning rate that distinguishes the LMS and the NLMS learning rules. In the LMS rule, the learning rate is chosen by trial and error, whereas in the NLMS rule, the learning rate is determined adaptively such that the square output error is minimized. The NLMS adaptation rule is self-normalizing.

In this paper we propose a learning algorithm of neuro-fuzzy geotomography based on the ART and NLMS to accelerate the convergence of learning phase. The learning algorithm has the property of iteratively attempting to match the line integral of the fuzzy model with the corresponding projection data. A constrained optimization problem, in which squared error between the line integral of the fuzzy model and corresponding projection data is used as an objective function, is solved to obtain corrections for elements of unknown parameter vector. We generalize the learning algorithm by introducing the Minkowski norm (the sum of s -th power of the correction) to the objective function instead of Euclidean norm (squared errors). In consequence of using the Minkowski norm more smooth distribution of a material parameter is obtained. We show some reconstruction results from the artificially generated data and collected field data.

2. Neurofuzzy Geotomography

In this section we review a neurofuzzy approach to computerized tomography[12]. Let A_{ik} denote the membership function of the k -th fuzzy rule in the domain of the i -th input variable x_i . The k -th rule is written as "If x_1 is A_{1k} and x_2 is A_{2k} , then y is w_k ." The conclusion part of the fuzzy reasoning rule that infers the output y is simplified as real number w_k . In the case of Gaussian membership function, A_{ik} is defined as

$$A_{ik}(x_i) = \exp\left(-\frac{(x_i - a_{ik})^2}{b_{ik}}\right) \quad (1)$$

where the parameters a_{ik} and b_{ik} ($i = 1, \dots, N$) are given for each k and are changed in the training procedure. The final output y is written as

$$y = f(x_1, x_2) = \sum_{k=1}^K \mu_k(x_1, x_2) \cdot w_k \quad (2)$$

where $\mu_k(x_1, x_2)$ is the compatibility degree of the premise part of the k -th fuzzy rule, which is defined as the product of membership function A_{ik}

$$\mu_k(x_1, x_2) = A_{1k}(x_1) \times A_{2k}(x_2) \quad (3)$$

In Fig.1 the line from A ($0, x_2^L$) to B ($1, x_2^R$) can be written as $x_2 = \alpha x_1 + x_2^L$ where $\alpha = x_2^R - x_2^L$. Let (x_1, x_2) be the

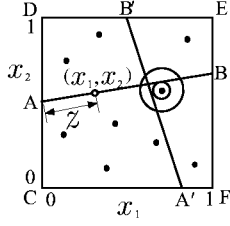


Figure 1. A region of probing.

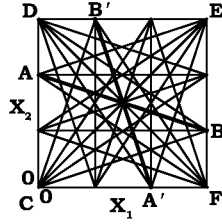


Figure 2. Propagation paths.

coordinate of a point on the line AB. Then the line integral of $f(x_1, x_2)$ along the path AB can be written as

$$\int_{-\infty}^{\infty} f(z) dz = \sum_{k=1}^K \zeta_k \sqrt{\pi b_k} \cdot w_k \quad (4)$$

where

$$\zeta_k = \exp \left(- \frac{(\alpha a_{1k} + x_2^L - a_{2k})^2}{(1 + \alpha^2) b_k} \right) \quad (5)$$

We here consider an application of neurofuzzy modeling to the geotomography. In Fig.2, a typical situation, the data collection system scans the rectangular region (CDEF) with some signal sources located in two boreholes CD and CF and some receivers in the opposite EF and DE. In Fig.2 typical ray paths (AB and A'B') are indicated for signal propagations. Let the line segment AB and A'B' be propagation paths L_p and L'_p respectively, and P be the number of propagation paths. We express the distribution of some material parameter such as attenuation rate by the fuzzy model $f(x_1, x_2)$ in Eq.(2), where (x_1, x_2) is the coordinate in the region. Assuming the straight line ray-optic model, the projection along a path L_p can be written as

$$I_p = \int_{L_p} f(x_1, x_2) dz, \quad (p = 1, 2, \dots, P) \quad (6)$$

Synthetic propagation data I_p^* can be measured along propagation paths. In a similar manner, the projections $I_p'^*$ along a path L'_p can be measured.

Let a cost function for the evaluation of the approximation be

$$E_1 = \frac{1}{2} \left\{ \sum_{p=1}^P (I_p - I_p^*)^2 + \sum_{p=1}^{P'} (I'_p - I_p'^*)^2 \right\} \quad (7)$$

and formulate a nonlinear mathematical programming problem to find the values of parameters a_{ik} , b_k and w_k in the neurofuzzy model, that minimize the cost function. If Gaussian functions are allocated inside the region, the value of Gaussian function $\mu_k(x_1, x_2)$ is almost 0 on the both boundaries of the region, then the line integral of the fuzzy model is approximated by Eq.(4). And in the case that Gaussian functions are allocated on the boundary of the region, since $\mu_k(x_1, x_2)$ is almost 0 on the opposite side of the region, the line integral is half of Eq.(4) approximately. The line integral I_p of $f(x_1, x_2)$ along the path L_p can be approximated as

$$I_p = \sum_{k=1}^K \zeta_{kp} \sqrt{\pi b_k} w_k + \frac{1}{2} \sum_{k=K+1}^{K+J} \zeta_{kp} \sqrt{\pi b_k} w_k \quad (8)$$

where K and J are the numbers of Gaussian functions that are allocated inside the region and on the boundaries of the region respectively.

It is hard to reconstruct detailed pictures of the spatial distribution from very small number of projection data. It is clearly an ill-posed problem because the information in the data is not sufficient to reconstruct the distribution in the region uniquely. Techniques that exploit smoothness constrains in order to transform an ill-posed problem into a well-posed one are well known under the term of regularization theory [8, 10]. Let a cost function for the regularization be as

$$E_2 = \frac{1}{2} \sum_{(x_1, x_2) \in W} \left\{ \left(\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} \right)^2 + \left(\frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} \right)^2 \right\} \quad (9)$$

where W is the set of data points chosen uniformly in the region. Let a total cost function be $E = E_1 + \eta E_2$, where η is positive weighting constant for regularization conditions, and formulate a nonlinear mathematical programming problem to find the values of parameters a_{ik} , b_k and w_k in the fuzzy model, which minimize the total cost function. The learning rules of a_{ik} , b_k and w_k are based on the steepest descent method(SDM).

3. Instantaneous Learning Algorithm Based on ART with Minkowski Norm

Generally speaking, it is an important and difficult problem to determine the appropriate learning rate. Usually learning rates in such learning algorithms that use gradient descent methods are chosen mainly by trial and error. However, there are two problems with the approach. If the learning rate is not small enough, the total sums of square errors can diverge to high values instead of decreasing. A very conservative estimate on the other hand will unnecessarily slow down the learning process. Algebraic Reconstruction Techniques(ART) algorithms used in geotomography have a simple procedure with repeated corrections to bring each projection of the estimate into agreement

with the corresponding measured projection. We propose a learning algorithm in which the learning rate changes adaptively using the concept of the ART.

Let a line integral of the fuzzy model along a propagation path be a function $y(\mathbf{w})$ of vector \mathbf{w} which consists of unknown parameter $a_{ik}, b_k, w_k, i = 1, 2, k = 1 \cdots K$. When $\Delta \mathbf{w}$ is a very small change of \mathbf{w} , $y(\mathbf{w})$ can be developed into a Taylor series with respect to \mathbf{w} as

$$y(\mathbf{w} + \Delta \mathbf{w}) = y(\mathbf{w}) + \nabla y(\mathbf{w})^T \Delta \mathbf{w} \quad (10)$$

The symbol ∇ indicates the differential operator $\partial y / \partial w_r$. To derive a learning algorithm based on the concept of the ART, let us consider the constrained optimization problem as

$$\min. \quad \sum_{r=1}^R (\Delta w_r)^s \quad (11)$$

$$s.t. \quad \sum_{r=1}^R \left| \frac{\partial y}{\partial w_r} \right| \Delta w_r = |\delta| \quad (12)$$

$$\Delta w_r \geq 0 \quad (13)$$

where, Δw_r is the correction for r -th element of unknown parameter vector \mathbf{w} , R is the total number of unknown parameters and δ is the difference between the measured value I_p^* and estimated value $y(\mathbf{w})$, that is, $\delta = y(\mathbf{w}) - I_p^*$. Note that s is a positive real number, then a constraint Eq.(13) is added as the non-negativity restrictions. The constrained optimization problem represented by Eqs.(11)–(13) is geometrically illustrated in Fig.3, where the 2-dimensional parameter space is shown. In Fig.3 the solid line represents a hyperplane ($y(\mathbf{w}) = I_p^*$) on which the constraint represented by Eq.(12) is satisfied. Let \mathbf{w}_n be a parameter vector after n learning iterations. A correction vector $\Delta \mathbf{w}$ is chosen such that a parameter vector \mathbf{w}_{n+1} after a correction can satisfy the constraint represented by Eq.(12), that is, the parameter vector \mathbf{w}_{n+1} must be on the hyperplane. The object of the constrained optimization problem is to minimize the sum of s -th power of corrections (Minkowski norm of corrections). When $s = 2.0$, a solution of the optimization problem equals the correction by the ART algorithm. Since the Euclidean distance of $\Delta \mathbf{w}$ is minimized in case of $s = 2.0$, $\Delta \mathbf{w}$ is orthogonal to the hyperplane. The parameter vector after a correction is denoted as \mathbf{w}'_{n+1} in Fig.3. When s approaches infinity, Eq.(11) is referred to as the minimax norm and the parameter vector after a correction is \mathbf{w}''_{n+1} in Fig.3. As s becomes large, all elements of correction vector $\Delta \mathbf{w}$ take almost same values.

We define the Lagrangian function as

$$L = \sum_{r=1}^R (\Delta w_r)^s + \lambda_0 \left(|\delta| - \sum_{r=1}^R \left| \frac{\partial y}{\partial w_r} \right| \Delta w_r \right) - \sum_{r=1}^R \lambda_{1r} \Delta w_r \quad (14)$$

where, λ_0, λ_{1r} are Lagrangian multipliers. If Δw_r is the local optimal solution, L, λ_0 and λ_{1r} satisfy the following Kuhn-Tucker conditions

$$\frac{\partial L}{\partial \Delta w_r} = s(\Delta w_r)^{s-1} - \lambda_0 \left| \frac{\partial y}{\partial w_r} \right| - \lambda_{1r} = 0 \quad (15)$$

$$\lambda_{1r} \Delta w_r = 0 \quad (16)$$

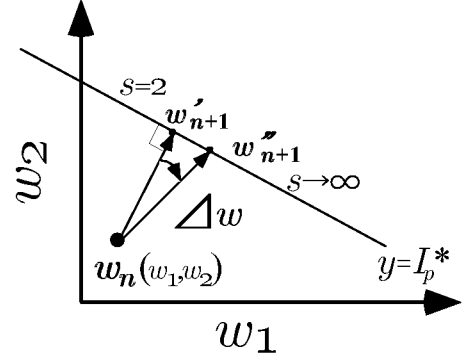


Figure 3. $\mathbf{w}, \Delta \mathbf{w}$ and hyperplane $y = I_p^*$ in the parameter space.

$$\Delta w_r \geq 0, \quad \lambda_0 \geq 0, \quad \lambda_{1r} \geq 0 \quad (17)$$

and

$$\frac{\partial L}{\partial \lambda_0} = |\delta| - \sum_{r=1}^R \left| \frac{\partial y}{\partial w_r} \right| \Delta w_r = 0 \quad (18)$$

for $r = 1, \dots, R$. Eqs.(15)~(17) lead to the solution

$$\Delta w_r = \left(\frac{\lambda_0}{s} \right)^{\frac{1}{s-1}} \left| \frac{\partial y}{\partial w_r} \right|^{\frac{1}{s-1}} \quad (19)$$

Combining Eq.(19) with Eq.(18) gives

$$|\delta| = \left(\frac{\lambda_0}{s} \right)^{\frac{1}{s-1}} \sum_{r=1}^R \left| \frac{\partial y}{\partial w_r} \right|^{\frac{s}{s-1}} \quad (20)$$

and we have

$$\Delta w_r = \frac{|\delta|}{\Sigma} \left| \frac{\partial y}{\partial w_r} \right|^{\frac{1}{s-1}} \quad (21)$$

where

$$\Sigma = \sum_{r=1}^R \left| \frac{\partial y}{\partial w_r} \right|^{\frac{s}{s-1}} \quad (22)$$

The solution Δw_r denotes absolute value of corrections, then we set the sign taking steepest descent direction into account as $-sgn(\delta \partial y / \partial w_r)$. The learning algorithm is given by

$$w_r^{NEW} = w_r^{OLD} - \tau sgn \left(\delta \frac{\partial y}{\partial w_r} \right) \Delta w_r \quad (23)$$

where, τ is a relaxation rate.

To demonstrate the effectiveness of Minkowski norm based on the ART, we carry out a numerical simulation. We generate a set of input output pair of data from a function denoted by dashed line in Fig.4. “•” represents the training data. The number of training data is 15 and each training data contains noise. The function is identified using the fuzzy model which consists of the 15 fuzzy rules. We set initial values of the parameters w_k, b_k to 0.0 and 0.01 respectively. a_k is uniformly spaced in unit interval [0,1]. Fig.4 shows the output of the fuzzy model (a) in case of

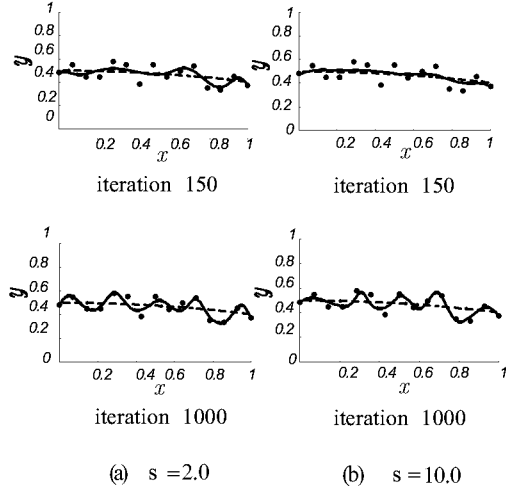


Figure 4. Effect of ART with Minchowski Norm

$s = 2.0$ (the original ART) and (b) in case of $s = 10.0$. (a) and (b) show the results after 150 iterations and 1000 iterations respectively. The outputs of fuzzy models are drawn by the solid curves. Comparing Fig.4(b) with (a), the output of the model obtained by the proposed method is more smooth at the early stage of learning iteration.

4. Computer Simulations

The performance of the reconstruction algorithm based on the ART is demonstrated using computer generated data. Fig.6(a) shows an artificially generated test distribution of material parameters such as attenuation rate or propagation velocity, which is drawn with 3-D graphics and contour curves. We locate four transmitters on a borehole CD and CF , and the four receivers on DE and EF respectively. We set 30 projection paths as shown in Fig.2. We used 16 fuzzy rules.

The proposed learning algorithm based on the ART is compared with the conventional learning algorithm SDM. In Fig.5 the value curves of the square sum of errors are plotted versus iteration number. The errors between the collected data and the output of the fuzzy model decrease more rapidly in the case of proposed ART than the SDM. We set $\tau = 0.1$ and $s = 2.0$ for the proposed algorithm and the learning rate to 0.001 for the SDM. The errors decreased to 0.019 and 0.107 at 1,000 iterations. Even after 10,000 iterations the error 0.031 remained by the SDM.

The reconstruction images are represented by both 3-D graphics and contour curves, and “o” represents a center of Gaussian function after learning. Gaussian functions were initially allocated uniformly inside the region and on the boundaries of the region. The reconstructed images by

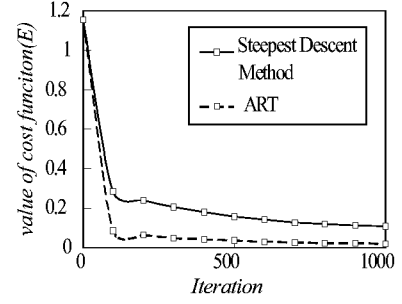


Figure 5. The value curves of cost function.

both learning algorithms are shown in Fig.6 where (a) the test distribution, (b) the result with the learning algorithm based on the ART at 10,000 iteration and (c) the result with the conventional learning algorithm at 100,000 iterations are shown respectively. Comparing (b) with (c) we cannot find much difference between them.

Next, we discuss the value of s in Eq.(11). In the reconstruction procedure we used the same training data generated from the test distribution shown in Fig.6(a). Three values of s were chosen. Figs.7 (a), (b) and (c) show the reconstruction images by the algorithms based on the ART with the parameters s being set to 3.0, 4.0 and 6.0 respectively. The squared error was smallest when $s = 3.0$. As shown in Fig.7, the output of the model(b) obtained in case of $s = 4.0$ is more smooth than (a).

5. Application to Field Data

The proposed algorithm was applied to the field data collected by cross-borehole seismic probing at a dam site in Mie Prefecture, Japan. A site plan with all observed ray paths is shown in Fig.8. Our aim is to determine whether the proposed method would be useful in detecting changes in a natural geologic structure. There are three vertical boreholes in Fig.8. The transmitters and receivers were spaced at 2.5m and 0.5m intervals respectively in each boreholes. In Fig.8 the transmitters and receivers are indicated by “•” and “o” respectively. The numbers of propagation paths were 269(No.1-No.2) and 240(No.2-No.3). Seismic waves were generated in the all boreholes and were measured in the all boreholes. As the received signals carry time of flight information for velocity measurements, measurements of the travel-time of seismic waves were employed to make an image of cross section. The travel-time is a line-integral of an inverse of velocity along each path, which is called as “slowness”. The distribution of slowness is obtained in case of using the travel-time as the training data.

We attempted to reconstruct the distribution of slowness from collected data of travel-time along the paths by the proposed algorithm. After that, we calculated the distribution of velocity which is the inverse of slowness. Fig.9

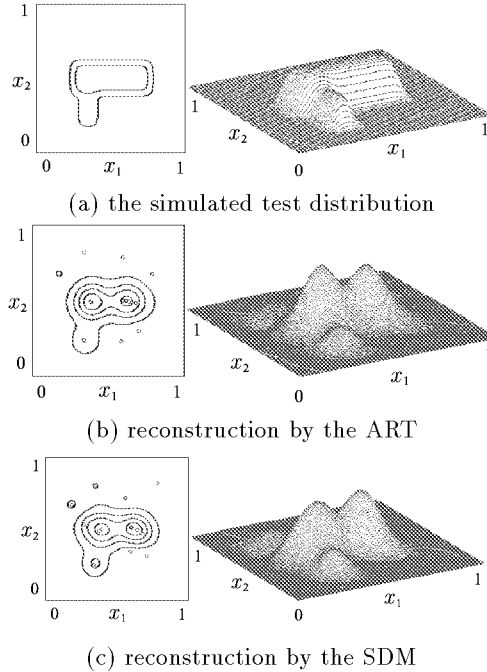


Figure 6. Results of the reconstruction. The number of propagation paths is 30. 16 Gaussian functions are used. (a) the simulated test distribution of a material parameter, (b) the output of $f(x_1, x_2)$ using learning algorithm based on ART, and (c) the output of $f(x_1, x_2)$ using steepest descent method.

shows the obtained distributions of the velocity represented with 3D-graphics and contour curves. The velocity contour interval is 1km/sec. The iteration number is set to 10,000. In Fig.9, the two different results in which s are set to 3.0 and 4.0 are illustrated. The results with $s = 3.0$ are shown in Fig.9(a). The results with $s = 4.0$ are shown in Fig.9(b). As shown in Figs.9(a) and (b), the output of the fuzzy model obtained with $s = 4.0$ is more smooth than with $s = 3.0$. The velocity distribution obtained by the proposed learning algorithm agrees with the knowledge of geological experts.

6. Conclusion

In this paper we have proposed a novel learning algorithm for the neurofuzzy geotomography based on the ART to accelerate the convergence of learning. A constrained optimization problem with the Minkowski norm is formulated and a learning algorithm using its solution is derived. It is demonstrated through some numerical simulations that smooth distribution of material parameters is obtained by

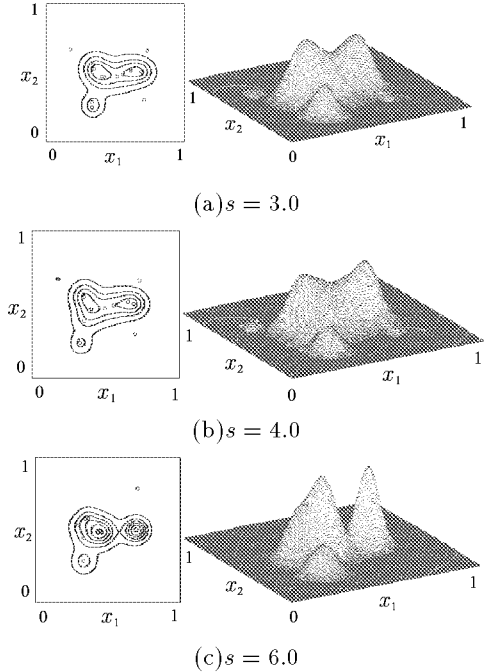


Figure 7. Results of the reconstruction with various values s , (a) $s = 3.0$, (b) $s = 4.0$ and (c) $s = 6.0$. 16 Gaussian functions are used.

the proposed algorithm. Furthermore, we applied the proposed algorithm to the field data and reconstructed the distribution of velocity of seismic wave at a dam site.

References

- [1] M.Brown and C.Harris, "Neurofuzzy Adaptive Modelling and Control", Prentice Hall, New York, 1994.
- [2] D.S.Broomhead and D.Lowe, "Multivariable Functional Interpolation and Adaptive Networks, Complex Systems", Vol.2, pp.321-355, 1988.
- [3] K.A.Dines and R.J.Lytle, "Computerized Geophysical Tomography", Proc. of IEEE, Vol.67, No.7, pp.1065-1073, 1979.
- [4] R.Gordon, R.Bender and G.T.Herman, "Algebraic Reconstruction Techniques(ART) for Three-dimensional Electron Microscopy and X-ray Photography", J. Theor. Biol. No.29, pp.471-481, 1970.
- [5] R.Gordon, "A Tutorial on ART", IEEE Trans. on Nuclear Science, Vol.NS-21, pp.78-93, 1974.
- [6] G.T.Herman, "Reconstruction of Binary Patterns from a Few Projection", Interna. Computing Symposium

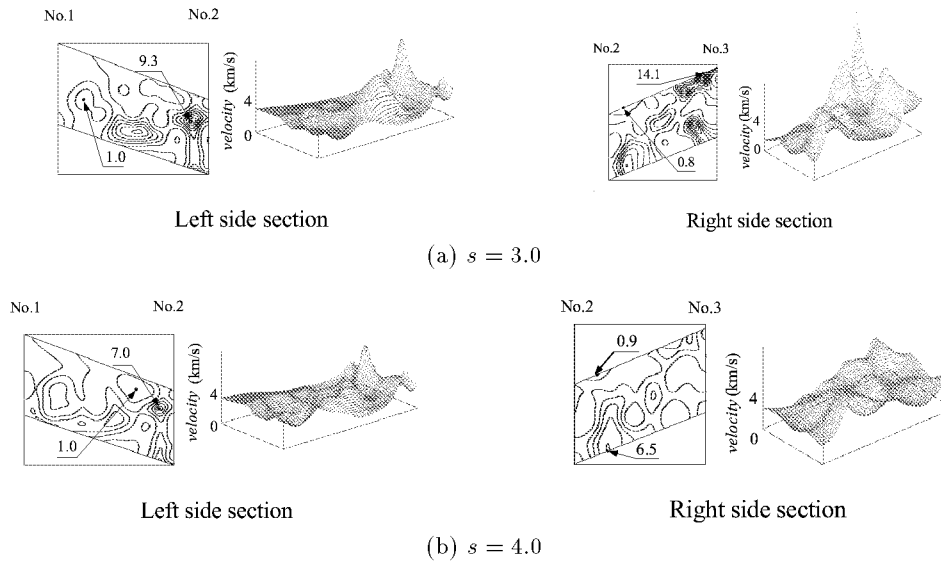


Figure 9. Image of cross section obtained by the proposed Geotomography. The velocity profile is shown. Spot values are given in km/sec. The velocity contour interval is 1km/sec.

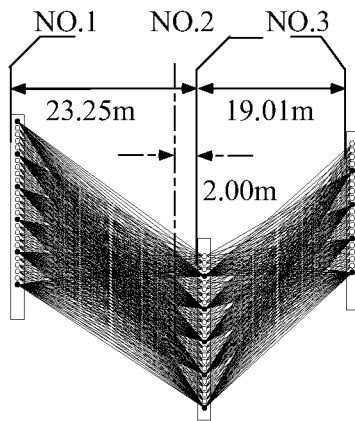


Figure 8. Cross-borehole sampling at a dam site using multiple locations for the transmitter and receiver probes

1973, ed.A.Guenther *et al.*, North Holland Publ. Co., pp.371–379,1974.

- [7] G.T.Herman, A.Lent and S.Rowland, “ART : Mathematics and Applications. A Report on the Mathematical Foundations and on the Applicability to Real Data of the Algebraic Reconstruction Techniques”, J. Theor. Biol. Vol.42, pp.1–32,1973.
- [8] T.Poggio and F.Girosi, “Regularization Algorithms for

Learning that Are Equivalent to Multilayer Networks”, Sciences, Vol.247, pp.978-982, 1990.

- [9] D.E.Rumelhart, J.L.McClelland and the PDP Research Group, “Parallel Distributed Processing”, Cambridge, MA: MIT Press, 1987.
- [10] M.Suzuki, K.Judd, K.Aihara and M.Kotani, “Approximation of the Logistic Mapping with Radial Basis Function Networks”, Transactions of IEICE, Vol.J76-A, No.8, pp.1177-1184, 1993.
- [11] D.W.Sweeney and T.M.Vest, “Reconstruction of Three-dimensional refractive Index Fields from Multidirectional Interferometric Data”, Applied Optics, Vol.12, No.11, pp.1649–2664,1973.
- [12] H.Tabuchi, T.Miyoshi, H.Ichihashi and K.Ohno, “Computerized Tomography with Radial Basis Functions Network”, Proc. of ICNN’95, pp.2258-2263, 1995.