Defect recognition in conductive materials by local magnetic-field measurement

T. Doi, S. Hayano, I. Marinova, N. Ishida, and Y. Saito

College of Engineering, Hosei University, Kajino, Koganei, Tokyo 184, Japan

Among the various nondestructive testing methods, the electric potential method requires relatively simple device and measurement. However, because of the high electrical contact resistance, sometimes it is difficult to measure the correct local electric potentials by direct contact. In order to overcome this difficulty, method is proposed which involves substituting the local magnetic field for electric potential measurements. The comparison with the conventional electric potential method demonstrates the usefulness this method, especially for materials with low resistivity.

I. INTRODUCTION

In order to prevent accidents in aircraft, iron bridges, and nuclear reactors, crack or defect recognition is of paramount importance. Various nondestructive testing methodologies, e.g., eddy current testing, x-ray computed tomography, ultrasonic imaging, and the electric potential method, have been exploited and utilized. Among these methods, the electric potential method requires relatively simple instruments and techniques.[1-3] However, because of the high electrical contact resistance and relatively low resistance of the target, sometimes it is difficult to measure the correct local electric potentials by direct contact. In order to remove this difficulty, we propose a new method which substitutes the local magnetic field for electric potential measurements. This makes it possible to implement the electric potential method without requiring direct contact to measure potentials. Most of the defect recognition problems are, in essence, reduced to solving inverse problems. Previously, the current distributions in the human heart and brain have been successfully estimated from the local magnetic fields by the sampled pattern matching (SPM) method. The SPM method has been exploited for solving the inverse source problems, and this is applicable to defect recognition problems, i.e., medium parameter identification problems.[4-7] We apply this SPM method with some modifications to our new electric potential method. The comparison with the conventional electric potential method demonstrates the usefulness of our new method, especially for materials with low resistivity.

II. PROJECTIVE SAMPLED PATTERN MATCHING METHOD

A. Key ideas

Let C, X, Y_s be the system matrix, potential, and input current vectors, respectively. Then, it is possible to write

$$CX = Y_s, \tag{1}$$

as a discretized electric current flowing system equation of a conductive material having a defect or crack. This system equation can be modified into

$$C_0 X = (C_0 - C)X + Y_s = Y + Y_s, \tag{2}$$

where C_0 is a system matrix without a defect or crack. This means that the potential vector X is composed of the two input vectors Y and Y_s. The vector Y is the equivalent field source vector

$$Y = (C_0 - C)X, \tag{3}$$

caused by the defect or crack. The vector Y_s is the externally impressed field source vector. Thus, the defect or crack recognition problems can be reduced into the equivalent source vector Y searching problems. Obviously, this equivalent source vector Y depends on the solution vector X. Further, this solution vector X is a function of the externally impressed field source vector Y_s. Thereby, the equivalent source vector Y can be expressed as a function of the impressed field source vector Y_s. That is,

$$Y = f(Y_s). \tag{4}$$

Introducing this functional relationship into our SPM method leads to the following advantages. First, this method requires a quite low CPU resource compared with those of the original SPM method, because the SPM process has to be carried out only for the vectors satisfying the relationship (4). Second, relatively accurate solutions can be expected, because one of the characteristics of the solution vector Y is known. Finally, changing the direction of vector Y_s reaches the correct solution vector Y similar to the computed tomography. This new method is called the projective sampled pattern matching (PSPM) method, because the functional relationship (4) is the directional dependence of the known vector Y_s. In the present article, we examine the characteristics of the PSPM method. As a result, we propose here the simple implementation techniques to improve the accuracy of PSPM method.

B. System equations

In Eq. (2), let us assume that C_0 is a square matrix of size m by m; X, Y, and Y_s are the mth-order column vectors, then taking the inverse matrix of C_0 and multiplying it to both sides of Eq. (2) yields

$$X = C_0^{-1} Y + C_0^{-1} Y_s = C_0^{-1} Y + X_0, \tag{5}$$

where $X_0 \, (= C_0^{-1} Y_s)$ is a potential vector without a defect or crack. Subtracting X_0 from Eq. (5), we have
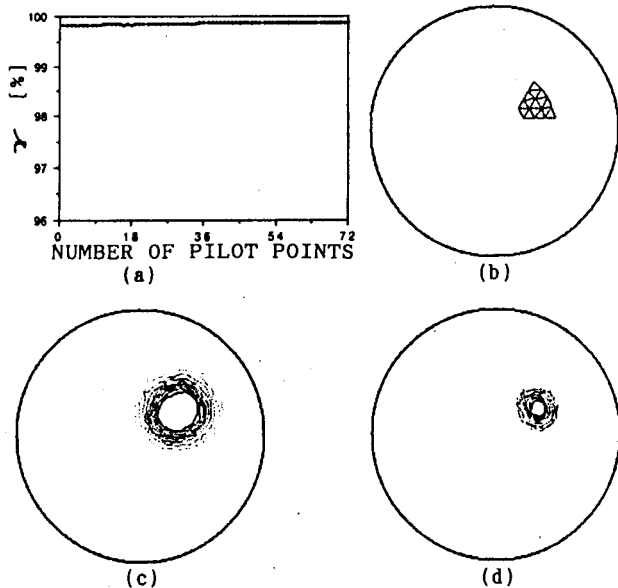
$$X_d = X - X_0 = C_0^{-1} Y. \tag{6}$$

FIG. 1. Single defect recognition by measuring the electric potential around a target area. (a) γ vs number of the pilot points in the processes of electric potential method, (b) exact solution, (c) recognized defect by measuring the electrical potentials (d) recognized defect by measuring the magnetic fields. The number of subdivisions $m = 1801$, and the number of measured points $n = 72$. The direction of the externally impressed current input vector Y_s was changed 72 times with 5° subdivision.
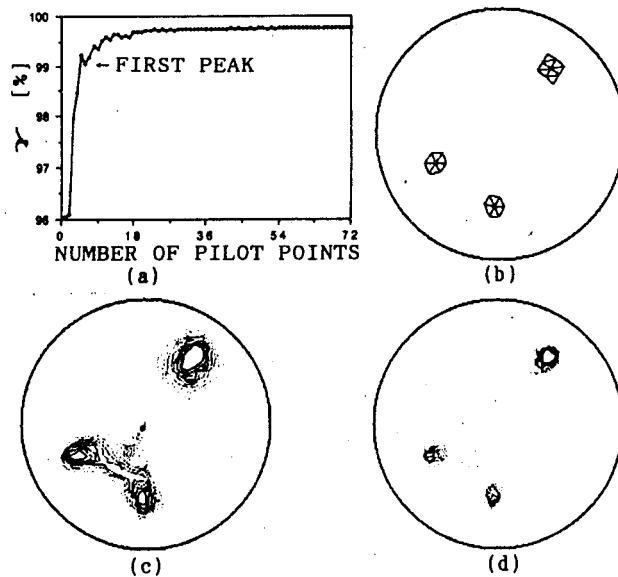
FIG. 2. Plural defects recognition by measuring the electric potential around a target area. (a) γ vs pilot points in the processes of electrical potential method, (b) exact solution, (c) recognized defect by measuring the electrical potentials (d) recognized defect by measuring the magnetic fields. The number of subdivisions $m = 1801$, and the number of measured points $n = 72$. The direction of the externally impressed current input vector Y_s was changed 72 times with 5° subdivision.

Generally, it is difficult to measure the entire difference vector X_d in Eq. (6) so that we can only obtain a part vector X_p of X_d. This yields the system equations of the defect or crack recognition problem. Namely, denoting n as a number of measured points, we have

$$X_p = DY = \sum_{i=1}^{m} y_i d_i, \qquad (7)$$

where D and X_p are the n by m partial matrix in C_0^{-1} and measured difference vector with order n, respectively. The y_i and d_i in Eq. (7) are the ith element of Y and ith column vector in D, respectively.

C. Sampled pattern matching SPM method

If a point h takes the maximum of

$$\gamma_i = X_p^T \cdot d_i / (|X_p||d_i|), \quad i = 1,2,...,m, \qquad (8a)$$

then h is the first pilot point. If a point g takes the maximum of

$$\gamma_{hj} = X_p^T \cdot (d_h + d_j) / (|X_p||d_h + d_j|), \quad j = 1,2,...,m, j \neq h, \qquad (8b)$$

then g is the second pilot point. A similar process to Eq. (8a) or (8b) is continued up to the first peak of γ.[5,6]

Thus, the pilot point solutions are given by the following.

If i is a pilot point, then:

$$y_i = 1,$$

if i is not a pilot point, then: $\qquad (9)$

$$y_i = 0.$$

Pilot point solutions of Eq. (7) mean that the magnitude of the solution is represented by the space occupying rate of unit input.

D. PSPM method

The SPM method is applicable to inverse problems where the solution vector Y is not a function of the known function Y_s,

$$Y \neq f(Y_s). \qquad (10)$$

However, if a relationship (4) is established, then the SPM processes are carried out only to the vectors d_i ($i = 1,2,...,m$) satisfying the relationship (4).

If medium parameter identification problems satisfy relationship (4), the SPM processes is continued up to the kth pilot point until the following condition is satisfied:

$$\frac{|X_p|}{|X_{0p}|} \leq \left| \sum_{j=1}^{k} d_j \right| \left| \sum_{i=1}^{m} d_i - \sum_{j=1}^{k} d_j \right|^{-1}, \qquad (11)$$

where the vector X_{0p} is composed of the measured potentials having a defect or crack.

This method is called the PSPM method. The PSPM method has the following advantages: (i) low CPU resource; (ii) relatively accurate solutions; (iii) changing the projective angles with Y_s and taking the average of entire solutions to reach the correct solution vector; (iv) the pilot point solutions are available if the known vector X_p encloses a target area.
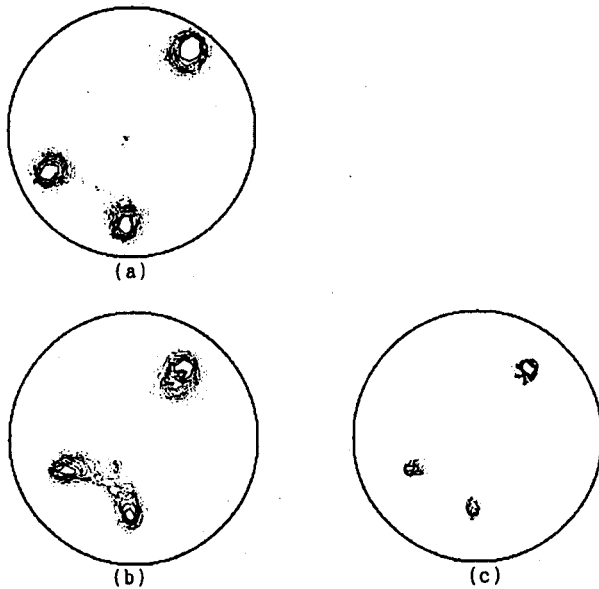
FIG. 3. Improved plural defects recognition: (a) Improvement by shrinking the target area, (b) and (c) show improved results by removing the pilot points up to the first peak for the electric potential and its modified methods, respectively.
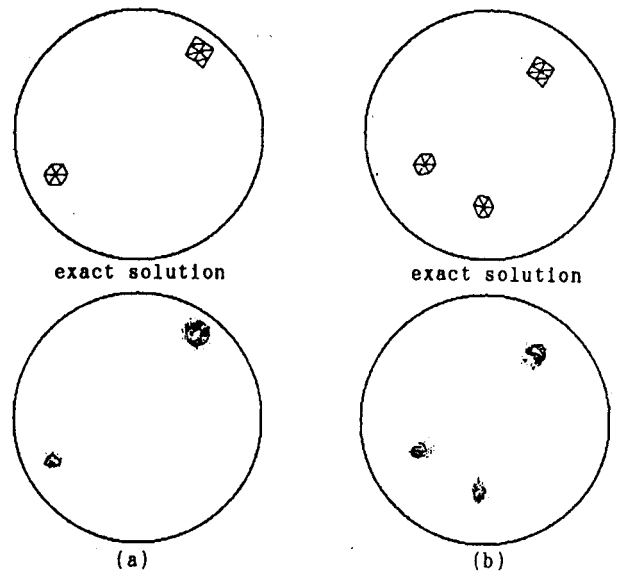

FIG. 4. The examples taking an average of the results obtained by electric potential and its modified methods (exact solution): (a) two defects, (b) three defects.

A curve obtained by plotting the pattern matching rate γ in Eq. (8a) or (8b) versus number of the pilot points reveals (i) if the curve continuously increases with addition of pilot points then the target area has a single defect or indistinguishable plural defects; (ii) if the curve has a deflection point then the object has plural defects. In this case, each of the plural defects can be distinguished by shrinking the target area or removing the pilot points up to the first peak value of γ. This is because the pilot points up to the first peak value of γ provide globally good pattern matching vectors common to the entire defects.

E. Examples

The two-dimensional electric potential method was chosen as the example. The electric potential method was carried out not only in the conventional way but also the modified way, which measured the magnetic fields instead of the electric potentials.

Figure 1(a) shows one of the γ versus number of pilot points curves in the processes of the electric potential method. Obviously, this curve continually increases with addition of pilot points, so that the target area has a single defect. In fact, results of PSPM method show the single defect as shown in Figs. 1(c) and 1(d).

Figure 2(a) shows one of the γ versus number of the pilot points curves. This curve has a deflection point, so that the target area must have the plural defects. In fact, the result of the PSPM method shows the plural defects as shown in Figs. 2(c) and 2(d).

Figure 3(a) shows improved plural defects recognition by shrinking the target area. Also, Figs. 3(b) and 3(c) show improved plural defects recognition by removing the pilot points up to the first peak of γ.

It must be noted here that the electric potential method always yields overestimated results in defect size. Conversely, the modified method yields underestimated results. Therefore, if it is possible to measure both electric potentials and magnetic field, the defect will stand midway between the results. Figure 4 shows examples of taking an average of both results.

Thus, the electric potential and its modified methods in combination with PSPM provide highly reliable results even if the plural defects are included in the target region.

III. CONCLUSION

As shown above, we have examined the PSPM method and proposed the methods of the improvement for the plural defects recognition problems. Examples have demonstrated that the plural defects recognition is improved by our method.

[1] S. Kubo, K. Ohji, and T. Sakagami, Int. J. Appl. Electromagn. Mater. 2, 81 (1991).
[2] Y. Iwamura and K. Miya, IEEE Trans. Magn. MAG-26, 618 (1990).
[3] T. Miyoshi and S. Nakano, Trans. Jpn. Soc. Mech. Eng. A 52, 1097 (1986).
[4] Y. Saito, E. Itagaki, and S. Hayano, J. Appl. Phys. 67, 5830 (1990).
[5] H. Saotome, K. Kitsuta, S. Hayano, and Y. Saito, Int. J. Appl. Electromagn. Mater. 3, 297 (1993).
[6] H. Saotome, T. Doi, S. Hayano, and Y. Saito, IEEE Trans. Magn. MAG-29, 1861 (1993).
[7] T. Doi, H. Saotome, S. Hayano, and Y. Saito, in Proceedings of the 2nd Japan-Hungary Joint Seminar, edited by T. Honma et al. (Hokkaido University Press, Sapporo, 1992), pp. 61–66.

Wavelet Solution of The Inverse Source Problems

Tatsuya Doi,   Seiji Hayano,   and   Yoshifuru Saito

College of Engineering,   Hosei University,   Kajino, Koganei, Tokyo 184, JAPAN

Abstract - Generally, the inverse source problem is reduced into solving an ill-posed system of equations. This article proposes a novel approach for the inverse source problem employing the wavelet analysis. The wavelet analysis has two distinguished abilities; one is the image data compression ability and the other is the spectrum resolution ability of the wave forms. Key idea is that the system matrix of the inverse source problems is regarded as a two-dimensional image data. The two-dimensional wavelet transform is applied to this system matrix. Finally, we can obtain an approximate inverse matrix of the system. A simple example demonstrates the validity of our approach.

I . INTRODUCTION

Inverse problems are classified into two major categories, i.e. one is the inverse parameter problem which evaluates the medium parameters by applying the electromagnetic fields to a target region and measuring its response; the other is the inverse source problem which evaluates the electromagnetic field sources from the locally measured electromagnetic fields. Generally, most of the inverse problems are reduced into solving the ill-posed system of equations.

Previously, we have proposed a method of solving for the inverse problems, and successfully applied to the biomagnetic fields as well as the nondestructive testing in metallic materials [1,2].

In the present article, we propose a novel approach utilizing the wavelet analysis. The wavelet analysis has been studied for the image data compression and analyzing the spectrum of image in informatics [3-6]. The wavelet analysis has two distinguished abilities; one is the image data compression ability and the other is the spectrum resolution ability of the wave forms. Key idea of our approach is that the system matrix of the inverse problems is regarded as a two-dimensional image data. The system matrix transformed into the wavelet spectrum space is composed of the two representative spectrums; one group has the larger absolute value, the other has nearly zero value. After collecting the spectrums having lager absolute value and building up the square matrix, an inverse of the square matrix is evaluated. Combining this local inverse matrix with the zero rectangular matrix, we apply the inverse wavelet transform to the resultant matrix. Thus, we have succeeded in obtaining an approximate inverse matrix of the inverse source problems.

II. DISCRETE WAVELET TRANSFORM

A. One-dimensional wavelet transform

In the present article, we employ the Haar's analyzing wavelets [3]. Let us consider a following linear transformation.

$$X' = CX, \tag{1}$$

where X is a data vector with order n; n must be a power of 2; and C is

$$C = \begin{bmatrix} c_0 & c_1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ c_1 & -c_0 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 0 & c_0 & c_1 & \cdot & \cdot & 0 & 0 \\ 0 & 0 & c_1 & -c_0 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & c_0 & c_1 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & c_1 & -c_0 \end{bmatrix} . \tag{2}$$

In equation (2), the first, third, fifth, and the other odd rows generate the components of data convolved with the coefficients c_0, c_1. This corresponds to a weighted integral operation. On the other side, the even rows generate the components of data convolved with the coefficients $c_1, -c_0$. This corresponds to a weighted differential operation [3,5].

In order to carry out an inverse linear transformation, the coefficients c_0, c_1 should be determined by a relationship:

$$C^T C = I , \tag{3}$$

where I is a n-th order unit matrix and a superscript T refers to the transpose of matrix C.

From equations (2) and (3), we have

$$c_0^2 + c_1^2 = 1. \tag{4}$$

Equation (4) has two unknowns c_0, c_1, but we have only one equation. To determine the coefficients c_0, c_1, generally, a following condition is considered:

$$c_0 - c_1 = 0. \tag{5}$$

From equation (4) and (5), we have

$$c_0 = \frac{1}{\sqrt{2}}, \qquad c_1 = \frac{1}{\sqrt{2}}. \tag{6}$$

The pair of coefficient c_0, c_1 in (6) is the Haar's analyzing wavelets, which are the same as the Daubechies's second order analyzing wavelets [3,4].

For simplicity, let us consider a data vector X with order 8:

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \end{bmatrix}^T. \tag{7}$$

Applying the transform matrix C_8 to (7)

$$X' = C_8 X = \begin{bmatrix} s_1 & d_1 & s_2 & d_2 & s_3 & d_3 & s_4 & d_4 \end{bmatrix}^T. \tag{8}$$

The elements in vector X' is sorted by using following matrix:

$$P_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

Thus, we have

$$P_8 X = P_8 C_8 X = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & d_1 & d_2 & d_3 & d_4 \end{bmatrix}^T. \tag{10}$$

Further transformation to the elements s_1, s_2, s_3, s_4 in (10) yields

$$W^{(2)}X = \begin{bmatrix} S_1 & S_2 & D_1 & D_2 & d_1 & d_2 & d_3 & d_4 \end{bmatrix}^T. \tag{11}$$

Similar transformation to S_1, S_2, D_1, D_2 in (11) yields

$$W^{(3)}X = \begin{bmatrix} S_1 & D_1 & D_2 & D_3 & d_1 & d_2 & d_3 & d_4 \end{bmatrix}^T. \tag{12}$$

The transformation matrix used in (11) and (12) are

$$W^{(2)} = (P_8' C_8')(P_8 C_8), \quad W^{(3)} = (P_8'' C_8'')(P_8' C_8')(P_8 C_8), \tag{13}$$

$$P_8' = \begin{bmatrix} P_4 & 0 \\ 0 & I_4 \end{bmatrix}, \quad C_8' = \begin{bmatrix} C_4 & 0 \\ 0 & I_4 \end{bmatrix}, \quad P_8'' = \begin{bmatrix} P_2 & 0 \\ 0 & I_6 \end{bmatrix}, \quad C_8'' = \begin{bmatrix} C_2 & 0 \\ 0 & I_6 \end{bmatrix}. \tag{14}$$

Equation (12) is the finally obtained wavelet spectrum. The elements S_1, D_1 in (12) are called the Mother Wavelet coefficients, and the others are called the wavelet coefficients at each level.

Inverse wavelet transform is carried out by

$$\begin{aligned} X &= [W^{(3)}]^T W^{(3)} X, \\ [W^{(3)}]^T &= [(P_8'' C_8'')(P_8' C_8')(P_8 C_8)]^T, \\ &= (P_8 C_8)^T (P_8' C_8')^T (P_8'' C_8'')^T, \\ &= C_8^T P_8^T (C_8')^T (P_8')^T (C_8'')^T (P_8'')^T. \end{aligned} \tag{15}$$

B. Two-dimensional wavelet transform

The discrete wavelet transform can be extended to the two dimensions [3]. Usually, two-dimensional wavelet transform is applied to a square matrix. In this article, two-dimensional wavelet transform is generalized to a rectangular matrix. The generalized two-dimensional wavelet transform is given by

$$M' = W_n M W_m^T, \tag{16}$$

where M' and M are the transformed (spectrum) matrix and original matrix with order n by m, respectively. W_n and W_m are the wavelet transform matrices with order n by n and m by m, respectively.

The inverse wavelet transform is carried out by the following equation:

$$M = W_m^T M' W_n. \tag{17}$$

III. THE INVERSE SOURCE PROBLEMS

A. Key idea

key idea is that the system matrix of the inverse source problems is regarded as one of the image data. The system matrix as an image data is transformed into a space of wavelet spectrum. The space of wavelet spectrum is composed of the two representative spectrums; one group has the larger absolute value, the other has nearly zero value. Collecting the spectrum having larger absolute value, we evaluate an inverse matrix of the system contracted to the non-singular size. Thus, the inverse wavelet transform yields an approximate inverse matrix of the system.

B. An example

An example of the inverse source problems is the estimation of current distribution on a film conductor from the locally measured magnetic fields. The estimation of

current distribution on a film is reduced into solving a following system equation

$$DX = Y, \tag{18a}$$

or

$$\begin{bmatrix} d_{11} & d_{12} & \cdots & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & \cdots & d_{2m} \\ \vdots & \vdots & & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & \cdots & d_{nm} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ \vdots \\ i_m \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_n \end{bmatrix}, \quad m \gg n, \tag{18b}$$

where D, X and Y are a system matrix determined by the Ampere's law, a current distribution vector to be estimated, and a measured magnetic field vector, respectively. Because of $m \gg n$, it is difficult to determine the vector X. The numbering of the current sources and measured magnetic fields have an effect on the quality of the wavelet transform [3,4]. In this paper, we employed a natural numbering.

Figure 1(a) shows a schematic diagram of the example. Our problem is that the current distributions on the film conductor is estimated from the locally measured magnetic fields. Figures 1(b) and 1(c) show the exact current distribution, and the measured magnetic fields, respectively. Figures 1(d) shows the system matrix D determined by the Ampere's law.

C. Approximate inverse matrix

In order to solve the system equation (18), we apply the discrete wavelet transform to the system matrix. Namely, the system matrix D is transformed into the wavelet spectrum D' by

$$D' = W_n D W_m^T. \tag{19}$$

Secondly, we take a square matrix S around the Mother wavelet coefficient out of the entire wavelet spectrum D'. After we take an inverse of the square matrix S, this inverse matrix S^{-1} is embedded to the matrix Z with order m by n.

$$D'^{-1}_{Appro} = S^{-1} \rightarrow Z. \tag{20}$$

Equation (20) means that the inverse matrix S^{-1} is embedded at the top square region of Z.

Finally, an approximate inverse matrix D^{-1}_{Appro} of the system is obtained by the two-dimensional inverse wavelet transform:

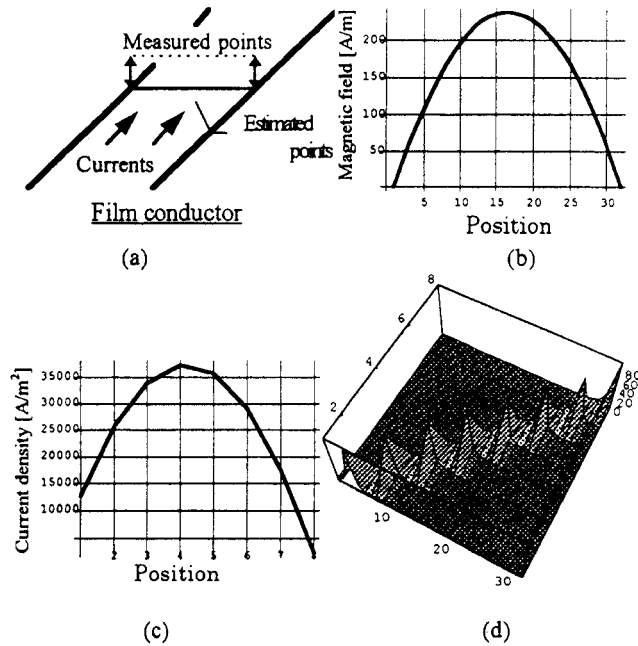$$D^{-1}_{Appro} = W_n^T D'^{-1}_{Appro} W_m. \tag{21}$$

Fig.1. (a) A schematic diagram, (b) an exact current distribution, (c) measured magnetic fields, and (d) the system matrix D represented as an image data.

Figures 2(a) and 2(b) show a two-dimensional wavelet spectrum D' of figure 1(d) and an approximate inverse matrix D^{-1}_{Appro}, respectively.

D. Validity of the approximate inverse matrix

Mathematical validity of the inverse matrix is generally carried out by means of the left- and right-inverse matrix checks. In this inverse source problem, the left-inverse matrix check $D^{-1}_{Appro} D$ is not equivalent to the right-inverse matrix check DD^{-1}_{Appro}, because the system matrix is a rectangular. When the left-inverse matrix check $D^{-1}_{Appro} D$ becomes

$$D^{-1}_{Appro} D = I_m, \tag{22}$$

the solution vector can be uniquely determined. Where I_m is an identity matrix with order m.

When the right-inverse matrix check DD^{-1}_{Appro} becomes

$$DD^{-1}_{Appro} = I_n, \tag{23}$$

The existence of solution vector can be confirmed. Where I_n is an identity matrix with order n.
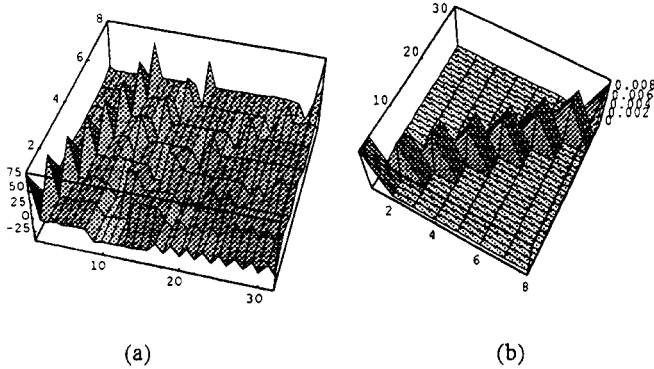
Fig. 2. (a) Two-dimensional wavelet spectrum D' of the system, and (b) an approximate inverse matrix D^{-1}_{Appro}.
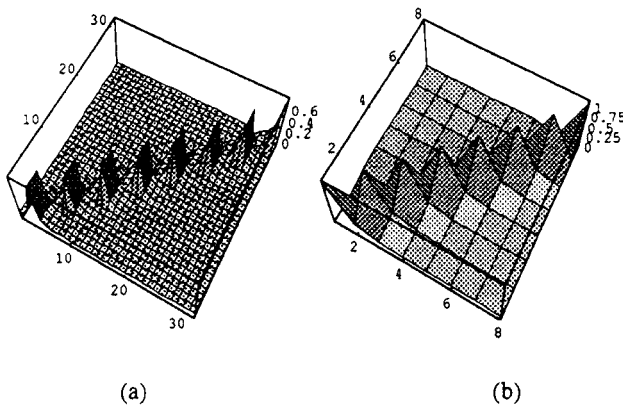

Fig.3. (a) The left-inverse matrix check $D^{-1}_{Appro}D$ and (b) the right-inverse matrix check DD^{-1}_{Appro}.

Thus, the left-inverse matrix check means the uniqueness of solution. The left-inverse matrix check shown in figure 3(a) is similar to the identity matrix I_n. This means that an approximate solution vector could be expected. Also, the right-inverse matrix check shown in figure 3(b) is the identity matrix I_n.This means that the existence of solution vector could be expected.

E. Wavelet solution

The current vector X in (18a) is given by

$$X = D^{-1}_{Appro} Y. \qquad (24)$$

Figures 4(a) and 4(b) show the estimated current distribution on the film conductor and reproduced magnetic fields, respectively.

Thus, we have succeeded in estimating the current distribution from the locally measured magnetic fields.
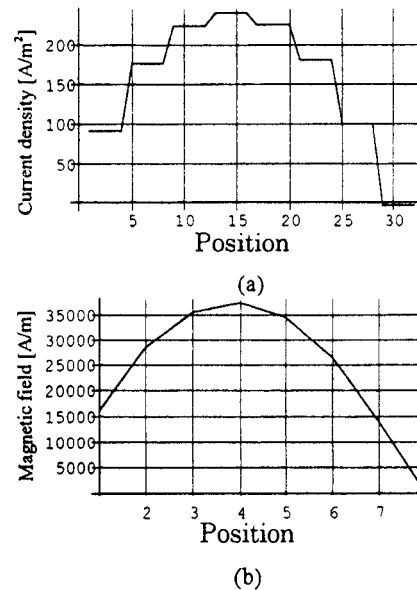

Fig.4. (a) The estimated current distribution on the film conductor, and (b) the reproduced magnetic fields from the estimated currents.

IV. CONCLUSION

In the present article, we have proposed an inverse approach employing the discrete wavelet transform. The two-dimensional wavelet analysis is applied to the rectangular system matrix as an image data. And the approximate inverse matrix of the system is obtained from a square part of the wavelet spectrum. Applying the inverse wavelet transform to the approximate inverse matrix in the wavelet spectrum space yields this approximate inverse matrix in the original space. Further, we have checked up the mathematical validity of the approximate inverse matrix.

The simple example concerning to the current estimation from the locally measured magnetic fields has demonstrated the validity of our approach.

REFERENCES

[1] H.Saotome, K.Kitsuta, S.Hayano, and Y.Saito, "A neural behavior estimation by the generalized correlative analysis," *IEEE Trans. on Magn.*, Vol. 29, No.2, pp. 1389-1394, Mar. 1993.

[2] T.Doi, S.Hayano, I.Marinova, N.Ishida, and Y.Saito, "Defect recognition in conductive materials by local magnetic-field measurement," *J.Appl.Phys.*, 75 (10), pp. 5907-5909 May 1994.

[3] D.E.Newland, "Random Vibrations, Spectral & Wavelet Analysis", Longman Scientific & Technical, 3rd Edit., 1993.

[4] I.Daubechies, "Orthonormal base of wavelets with finite support - Connection with discrete filter," *Comm. Pure and Appl. Math.* , 41, pp. 909-996, 1988.

[5] S.G.Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.11, No. 7, pp. 674-693, July 1989.

[6] N.V.Thakor, G.Xin-Rong, S.Yi-Chun, and D.F.Hanley, "Multiresolution wavelet analysis of evoked potentials," *IEEE Trans. on Biomed. Eng.*,40, pp. 1085-1094, 1993.

Wavelet Solution of The Inverse Parameter Problems

Tatsuya Doi,   Seiji Hayano,   and   Yoshifuru Saito

College of Engineering,   Hosei University,   Kajino, Koganei, Tokyo 184, JAPAN

Abstract - Previously, we have proposed a method of solving inverse problems, and successfully applied the method to biomagnetic fields as well as the nondestructive testing in metallic materials. In the present article, we propose a novel inverse approach for the parameter determination problems employing wavelet analysis. A simple example of parameter determination demonstrates the validity of our wavelet approach.

I . INTRODUCTION

Inverse problems are classified into two major categories, i.e. one is the inverse parameter problem; the other is the inverse source problem. For the inverse parameter problem, it is possible to obtain a unique solution if the fields are measured ideally; such as medium parameter identification in human body employing the computed tomography (CT). However, most of the inverse problems are generally reduced to solving a system equation for which it is difficult to obtain a unique solution. In order to overcome this difficulty, we have previously proposed a method of solving the inverse problems, and successfully applied it to biomagnetic fields as well as to nondestructive testing in metallic materials [1,2].

On the other hand, the wavelet analysis has been studied for image data compression and analyzing the spectrum of image in informatics [3-6].

In the present article, we propose a novel approach for the inverse parameter problems employing wavelet analysis. The key idea is that a system matrix of the inverse problems is regarded as two-dimensional image data. The two-dimensional wavelet transform is applied to this system matrix. An approximate inverse matrix of the system is obtained from the wavelet spectrum. We here consider a test example in which the relationship between input and output is evaluated from given input and output data. As a result, the example demonstrates the validity of our wavelet approach.

II . DISCRETE WAVELET TRANSFORM

A. One-dimensional wavelet transform

In the present paper, we employ Haar's analyzing wavelets [3]. Let us consider a following linear transformation

$$\mathbf{X}' = C\mathbf{X}, \qquad (1)$$

where $\mathbf{X}$ is a data vector with order n; n must be a power of 2; and C is

$$C = \begin{bmatrix} c_0 & c_1 & 0 & 0 & \cdot\cdot & 0 & 0 \\ c_1 & -c_0 & 0 & 0 & \cdot\cdot & 0 & 0 \\ 0 & 0 & c_0 & c_1 & \cdot\cdot & 0 & 0 \\ 0 & 0 & c_1 & -c_0 & \cdot\cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ 0 & 0 & 0 & 0 & \cdot\cdot & c_0 & c_1 \\ 0 & 0 & 0 & 0 & \cdot\cdot & c_1 & -c_0 \end{bmatrix} \cdot \qquad (2)$$

In equation (2), the first, third, fifth, and the other odd rows generate the components of data convolved with the coefficients c_0, c_1. This corresponds to a weighted integral operation. On the other hand, the even rows generate the components of data convolved with the coefficients $c_1, -c_0$. This corresponds to a weighted differential operation [3,5].

In order to carry out an inverse transformation, the coefficients c_0, c_1 should be determined by a relationship:

$$C^T C = I, \qquad (3)$$

where I is a n-th order unit matrix and a superscript T refers to the transpose of matrix C.

From equations (2) and (3), we have

$$c_0^2 + c_1^2 = 1. \qquad (4)$$

Equation (4) has two unknowns c_0, c_1, but we have only one equation. To determine the coefficients c_0, c_1, generally, a following conditions is considered:

$$c_0 - c_1 = 0. \qquad (5)$$

From equations (4) and (5), we have

$$c_0 = \frac{1}{\sqrt{2}}, \qquad c_1 = \frac{1}{\sqrt{2}}. \qquad (6)$$

The pair of coefficient c_0, c_1 in (6) is Haar's analyzing wavelets, which are equivalent to Daubechies's second order

analyzing wavelets [3,4].

For simplicity, let us consider a data vector $\mathbf{X}$ with order 8:

$$\mathbf{X} = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7 \quad x_8]^T. \tag{7}$$

Applying the transform matrix C_8 to (7) yields

$$\mathbf{X}' = C_8\mathbf{X} = [s_1 \quad d_1 \quad s_2 \quad d_2 \quad s_3 \quad d_3 \quad s_4 \quad d_4]^T. \tag{8}$$

The elements in vector $\mathbf{X}'$ are sorted by using the following matrix:

$$P_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

Thus, we have

$$P_8\mathbf{X} = P_8 C_8\mathbf{X} = [s_1 \quad s_2 \quad s_3 \quad s_4 \quad d_1 \quad d_2 \quad d_3 \quad d_4]^T. \tag{10}$$

Further transformation to the elements s_1, s_2, s_3, s_4 in (10) yields

$$W^{(2)}\mathbf{X} = [S_1 \quad S_2 \quad D_1 \quad D_2 \quad d_1 \quad d_2 \quad d_3 \quad d_4]^T, \tag{11}$$

where

$$W^{(2)} = (P_8'C_8')(P_8 C_8), \quad P_8' = \begin{bmatrix} P_4 & 0 \\ 0 & I_4 \end{bmatrix}, \quad C_8' = \begin{bmatrix} C_4 & 0 \\ 0 & I_4 \end{bmatrix}. \tag{12}$$

Similar transformation to S_1, S_2, D_1, D_2 in (11) yields

$$W^{(3)}\mathbf{X} = [S_1 \quad D_1 \quad D_2 \quad D_2 \quad d_1 \quad d_2 \quad d_3 \quad d_4]^T. \tag{13}$$

where

$$W^{(3)} = (P_8''C_8'')(P_8'C_8')(P_8 C_8),$$
$$P_8'' = \begin{bmatrix} P_2 & 0 \\ 0 & I_6 \end{bmatrix}, \quad C_8'' = \begin{bmatrix} C_2 & 0 \\ 0 & I_6 \end{bmatrix}. \tag{14}$$

The wavelet transform of one-dimensional data with order 8 is finally given by $3 \, (= Log_2 8)$ steps of linear transformation. Equation (13) is the finally obtained wavelet spectrum. The elements S_1, D_1 in (13) are called the Mother Wavelet coefficients, and the others are called the wavelet coefficients at each level.

Inverse wavelet transform is carried out by

$$\mathbf{X} = [W^{(3)}]^T (W^{(3)}\mathbf{X}),$$
$$[W^{(3)}]^T = [(P_8''C_8'')(P_8'C_8')(P_8 C_8)]^T,$$
$$= (P_8 C_8)^T (P_8'C_8')^T (P_8''C_8'')^T, \tag{15}$$
$$= C_8^T P_8^T (C_8')^T (P_8')^T (C_8'')^T (P_8'')^T.$$

B. Two-dimensional wavelet transform

The discrete wavelet transform can be extended to two dimensions [3]. Usually, two-dimensional wavelet transform is applied to a square matrix. In this article, two-dimensional wavelet transform is generalized to a rectangular matrix. The generalized two-dimensional wavelet transform is given by

$$M' = W_n M W_m^T, \tag{16}$$

where M' and M are the transformed (spectrum) matrix and original matrix with order n by m, respectively. W_n and W_m are the wavelet transform matrices with order n by n and m by m, respectively.

The inverse wavelet transform is carried out by the following equation:

$$M = W_m^T M' W_n. \tag{17}$$

III. THE INVERSE PARAMETER PROBLEMS

A. Wavelet approach

The key idea is that the system matrix is regarded as one of the image data. The system matrix as an image data is transformed into a space of wavelet spectrum.

Let us consider an inverse parameter problem. The system $\mathbf{X} = C\mathbf{Y}$ can be modified by exchanging the elements in the vector $\mathbf{Y}$ and matrix C, viz.

$$\mathbf{X} = Y\mathbf{C}, \tag{18}$$

where a matrix Y and vector $\mathbf{C}$ are the system matrix composed of the elements in $\mathbf{Y}$ and parameter vector to be determined, respectively. In order to solve for (18), we apply the two-dimensional discrete wavelet transform to (18). The system matrix Y is transformed by

$$Y' = W_n Y W_m^T. \tag{19}$$

From the result of (19), it is found that the spectrum matrix can be classified into two major groups. One group

takes the large absolute value, and the other takes the smaller absolute value. We take a square matrix S around the Mother wavelet coefficient out of the entire wavelet spectrum Y'. Generally, the square matrix S around the Mother wavelet coefficients have large values. After taking of the inverse matrix of S, we embed this inverse matrix into a zero matrix Z with order m by n.

$$Y'^{-1}_{Appro} = S^{-1} \rightarrow Z. \qquad (20)$$

Equation (20) means that the inverse matrix S^{-1} is embedded at the top square region of Z.

The approximate inverse matrix Y^{-1}_{Appro} of the system is obtained by the two-dimensional inverse wavelet transform:

$$Y^{-1}_{Appro} = W_n^T Y'^{-1}_{Appro} W_m. \qquad (21)$$

Finally, the parameter vector of system C is given by

$$C = Y^{-1}_{Appro} X. \qquad (22)$$

Thus, the parameter vector C can be obtained from the known input Y and output X.

B. An example

Let us consider an example of parameter identification problems. For this example, the current and magnetic field distribution are known vectors, but the relationship between them is unknown. This example is reduced to solving for the following system equation

$$X = YC, \qquad (23a)$$

or

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 & \cdots & y_n & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdots & 0 & y_1 & \cdots & y_n & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & y_1 & \cdots & y_n \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ \vdots \\ \vdots \\ c_{mn} \end{bmatrix}, \qquad (23b)$$

where C, X and Y are a vector of system parameter to be determined, an output vector, and the system matrix composed of the input current, respectively.

Figure 1 shows an example of a parameter identification problem from both input currents and output magnetic field vectors. Actually, exact parameter of the vector C in (23b) are determined by the Ampere's law. We verify that the exact parameter can be identified by the wavelet approach.
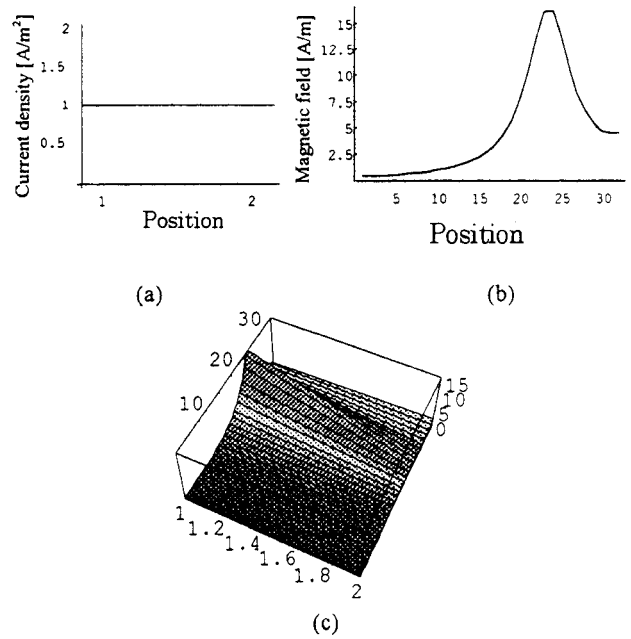

(a)   (b)

(c)

Fig.1. (a) An input current vector Y, (b) an output magnetic field vector X, and (c) the system matrix.
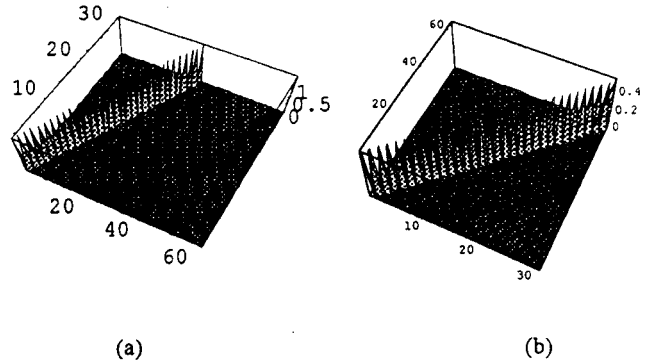

(a)   (b)

Fig. 2. (a) Two-dimensional wavelet spectrum Y' of the system, and (b) an approximate inverse matrix Y^{-1}_{Appro}.

Figures 1(a), 1(b) and 1(c) show an input current distribution, an output magnetic field distribution, and the system matrix of this parameter identification problem in (23b), respectively.

Figures 2(a) and 2(b) show a two-dimensional wavelet spectrum Y' of the system in figure 1(c) and an approximate inverse matrix Y^{-1}_{Appro} of the system, respectively.

Finally, the parameter vector C in (23) is given by

$$C = Y^{-1}_{Appro} X. \qquad (24)$$

Figures 3(a) and 3(b) show the determined parameter of the system, and reproduced magnetic field distribution, respectively. The result in 3(a) coincides with those of the Ampere's law. Actually, the parameter of the system is determined by the Ampere's law. Thus, we have succeeded
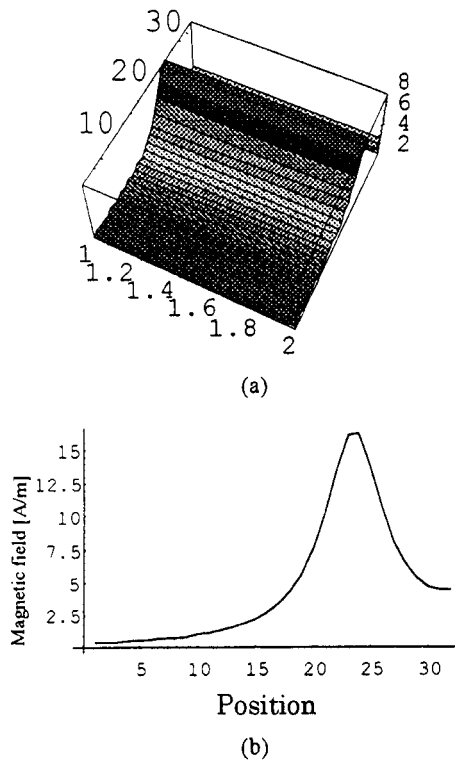
(a)

(b)

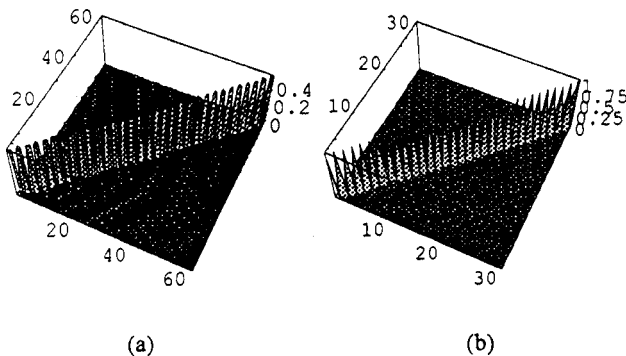Fig.3. (a) The decided system matrix, and (b) the reproduced magnetic field distribution by the estimation system matrix.

(a)   (b)

Fig.4. (a) The left-inverse matrix check $Y^{-1}_{Appro}Y$ and (b) the right-inverse matrix check YY^{-1}_{Appro}.

in estimating the parameter of the system from both the input and output vectors.

C. Validity of the approximate inverse matrix

Mathematical validity of the inverse matrix is generally carried out by means of the left- and right-inverse matrix checks. In this inverse parameter problem, the left-inverse matrix check $Y^{-1}_{Appro}Y$ is not equivalent to the right-inverse matrix check YY^{-1}_{Appro}, because the system matrix is

rectangular. When the left-inverse matrix check $Y^{-1}_{Appro}Y$ becomes

$$Y^{-1}_{Appro}Y = I_m,\qquad(25)$$

the solution vector can be uniquely determined, where I_m is an identity matrix with order m.

When the right-inverse matrix check YY^{-1}_{Appro} becomes

$$YY^{-1}_{Appro} = I_n,\qquad(26)$$

the existence of solution vector can be confirmed, where I_n is an identity matrix with order n.

Thus, the left-inverse matrix check means the uniqueness of solution. The left-inverse matrix check shown in figure 4(a) is similar to the identity matrix I_m. This means that an approximate solution vector could be expected. Also, the right-inverse matrix check shown in figure 4(b) is the identity matrix I_n. This means that the existence of solution vector could be expected.

IV. CONCLUSION

In the present paper, we have proposed a novel approach for the inverse parameter problems employing the wavelet analysis. The wavelet analysis is applied to the system matrix of the inverse parameter problems. The results reveal that our wavelet approach is possible to get an approximate inverse matrix of the system. A simple example has demonstrated the validity of our approach.

REFERENCES

[1] H.Saotome, K.Kitsuta, S.Hayano, and Y.Saito, "A neural behavior estimation by the generalized correlative analysis," IEEE Trans. on Magn., Vol. 29, No.2, pp. 1389-1394, Mar. 1993.

[2] T.Doi, S.Hayano, I.Marinova, N.Ishida, and Y.Saito, "Defect recognition in conductive materials by local magnetic-field measurement," J.Appl.Phys., 75 (10), pp. 5907-5909 May 1994.

[3] D.E.Newland, "Random Vibrations, Spectral & Wavelet Analysis", Longman Scientific & Technical, 3rd Edit., 1993.

[4] I.Daubechies, "Orthonormal base of wavelets with finite support - Connection with discrete filter," Comm. Pure and Appl. Math. , 41, pp. 909-996, 1988.

[5] S.G.Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Machine Intell., Vol.11, No. 7, pp. 674-693, July 1989.

[6] N.V.Thakor, G.Xin-Rong, S.Yi-Chun, and D.F.Hanley, "Multiresolution wavelet analysis of evoked potentials," IEEE Trans. on Biomed. Eng.,40, pp. 1085-1094, 1993.

Dose optimization of proton and heavy ion therapy using generalized sampled pattern matching

Kiyoshi Yoda†‡, Yoshifuru Saito§ and Hidenobu Sakamoto‖

† Mitsubishi Electric Corporation, Advanced Technology R&D Center, 8-1-1 Tsukaguchi-Hommachi, Amagasaki 661 Japan
§ Hosei University, College of Engineering, Department of Electronics & Electrical Engineering, 3-7-2 Kajino, Koganei, Tokyo 184 Japan
‖ Mitsubishi Electric Corporation, Communications Equipment Works 8-1-1 Tsukaguchi-Hommachi, Amagasaki 661 Japan

Abstract. We have proposed a new dose optimization method for proton and heavy ion therapy using generalized sampled pattern matching, where an optimal beam weight distribution for scanning is obtained as a solution. Using water phantom models, one-dimensional lateral and depth dose distributions were separately optimized, each resulting in a uniform dose distribution within a target region and minimum dose fall-off to minimize undesired irradiation onto neighbouring tissues. Subsequently, we have applied the technique to broad beam three-dimensional proton therapy, leading to a homogeneous dose distribution inside a target and minimized distal and lateral dose fall-off for most convex tumour shapes.

1. Introduction

Optimization of proton or heavy ion therapy ideally requires a uniform dose distribution within a target volume, a zero dose level in critical organs, and a minimum dose level in other normal tissues.

Several optimization techniques have been reported using a least-squares method (Pedroni 1995) or an iterative deconvolution scheme (Lind 1988, Brahme *et al* 1989) to obtain beam weight solutions for a given dose distribution. The unconstrained least-squares method may lead to unphysical negative beam weight solutions. The solution also depends on initial values, and it is not always easy to guess good initial values. This initial value problem may be overcome by employing random methods such as simulated annealing; however, this better strategy still has the negative solution problem. Meanwhile, the iterative deconvolution method is not directly applicable for inhomogeneous media, because the formulation is based on convolution of a spatially invariant energy deposition kernel.

We have proposed a new dose optimization method for proton and heavy ion therapy using generalized sampled pattern matching (Saito and Yoda 1996), where the resulting beam weights are always non-negative as for the previous iterative deconvolution scheme, and moreover a heterogeneous medium may be considered once we calculate spatially varying energy deposition kernels using Monte Carlo methods (Goitein and Sisterson 1978) or other

‡ E-mail address: yoda@ele.crl.melco.co.jp

approximated formulations such as semi-infinite slab models (Larsson 1987, Russell *et al* 1995). In this report, we show preliminary dose optimization results for homogeneous water phantom models.

2. Principle of generalized sampled pattern matching

Generalized sampled pattern matching (GSPM) is a generalization of an inverse problem solver, sampled pattern matching (Saito *et al* 1990, Saotome *et al* 1993, 1995, Yoda *et al* 1997). The original sampled pattern matching (SPM) provides only a one-bit (0 or 1) solution while the GSPM gives a multiple-bit solution. Here, we describe the GSPM principle using the terminology of three-dimensional proton dose optimization.

The basic equation we want to solve is given as follows:

$$D(r) = \sum_i N_i H_i(r) \tag{1}$$

where r is a three-dimensional spatial position (x, y, z), $D(r)$ is the desired dose distribution inside the medium, $H_i(r)$ is a dose distribution given by a monoenergetic Gaussian beam often called an energy deposition kernel and N_i is a beam weight or the number of incident particles. In GSPM, the beam weight solution N_is are approximated by quantized discrete values, and therefore the solution process is based on a discrete search algorithm. The GSPM involves two important aspects: unique objective function and a unique solving procedure.

The objective function for the GSPM algorithm is cosine of the vector angle between a desired dose distribution vector $U = (u_1, u_2, \ldots, u_m)$ and a theoretically calculated dose distribution vector $V = (v_1, v_2, \ldots, v_m)$ where m denotes the number of dose sampling points. In other words, each u_i ($i = 1, \ldots, m$) corresponds to a desired dose $D(r_i)$ in equation (1) that is predetermined, whereas each v_i ($i = 1, \ldots, m$) corresponds to a theoretically calculated dose given by the right-hand side of equation (1). The cosine of the angle θ of the above two vectors U and V is given by the inner product of the corresponding normalized vectors as follows:

$$\cos\theta = \frac{U \cdot V}{|U||V|} \tag{2}$$

$\cos\theta$ is called the pattern matching index. If the two vectors U and V lie in a same direction, then the value of $\cos\theta$ is 1.0, giving the best pattern matching regardless of the vector amplitudes $|U|$ and $|V|$. Conventional optimization methods always try to minimize the squared sum of the difference between the desired values and the calculated results; consequently, we have to consider absolute dose levels from the beginning possibly leading to undesirable local minimum solutions. In contrast, the GSPM finds a normalized beam weight distribution as a solution thereby highly reducing such local minimum problems. The absolute beam weight or irradiation dose can easily be obtained by dividing the desired absolute dose by the calculated relative dose after the optimization.

The optimization procedure resembles piling building blocks on the floor one-by-one. In this case, we add the individual beams, $H_i(r)$ (defined in equation (1)), together in an incremental fashion. Each beam is added with an incremental or unit beam weight (of say 0.1), and the same beam may be added with this incremental weight several times in the procedure if necessary. The first step is to choose one beam from the set of beams $\{H_i(r)\}$ that maximizes $\cos\theta$. For example, consider a one-dimensional problem of finding the weighting factors in the superposition of monoenergetic Bragg peaks to obtain a spread-out Bragg peak (i.e. the one-dimensional version of equation (1)). The beam that maximizes

$\cos\theta$ is the most energetic beam (i.e. the beam with the longest range) because for all other beams there will be some terms in the sum $U \cdot V = \sum u_i v_i$ which are zero. The beam with the longest range is the 'most important' beam in the superposition, and it is the one that is recognized first by the pattern matching index. The optimization process continues by adding in beams incrementally such that the pattern matching index, $\cos\theta$, is maximized for each addition step. This is repeatedly performed until all possible beams are included. Finally, we obtain an optimal weight distribution by finding the highest pattern matching index for all the combinations of beams considered.

Figure 1 shows the process of finding the solution. For simplicity a one-dimensional example is shown. The GSPM yields a solution after placing all the possible beam weights as shown in figure 1. The optimal weight distribution is obtained by selecting the distribution with the highest value of $\cos\theta$. This is represented by one of the intermediate diagrams in figure 1. In this figure the incremental or unit beam weight is chosen to be 0.1 and the maximum weight is set at 1.0 thereby providing a 10-step dynamic range for the optimal weight solution. For more precision, we simply choose a smaller unit weight such as 0.01 while maintaining the maximum weight unchanged. All the calculations were initially done by *Mathematica* (Wolfram 1991). Subsequently, the GSPM routine was coded by C and called from inside *Mathematica*.
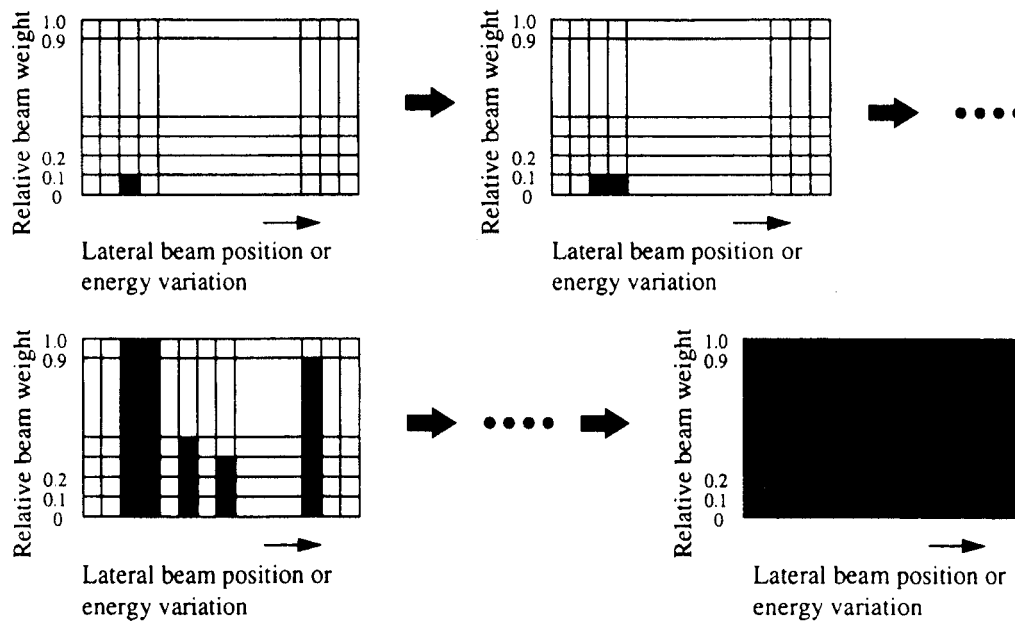

Figure 1. Diagram of solution search processes of generalized sampled pattern matching (GSPM). The solution is given as the beam weight distribution having the highest pattern matching index, $\cos\theta$.

3. Results

3.1. One-dimensional lateral dose optimization

The lateral dose distribution $D(x, z_0)$ at depth z_0 is given as follows:

$$D(x, z_0) = \sum_i N_i \frac{\exp[-(x - i\Delta x)^2/\sigma^2]}{\pi\sigma^2} S(E, z_0) \tag{3}$$
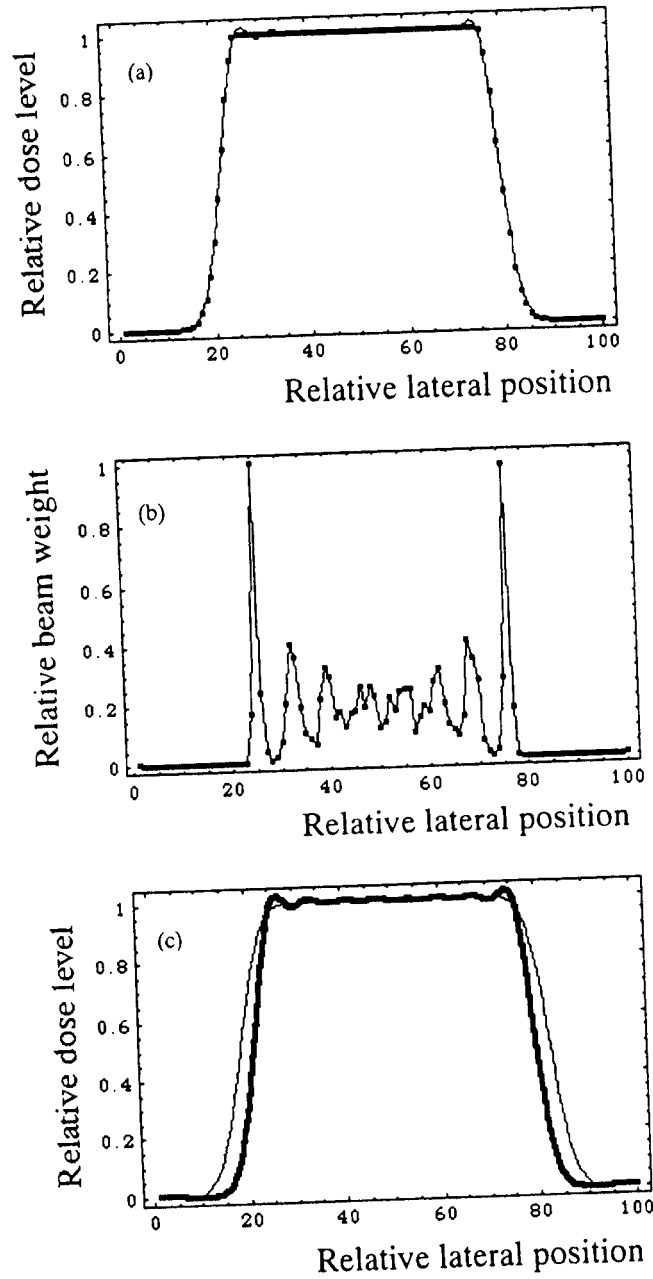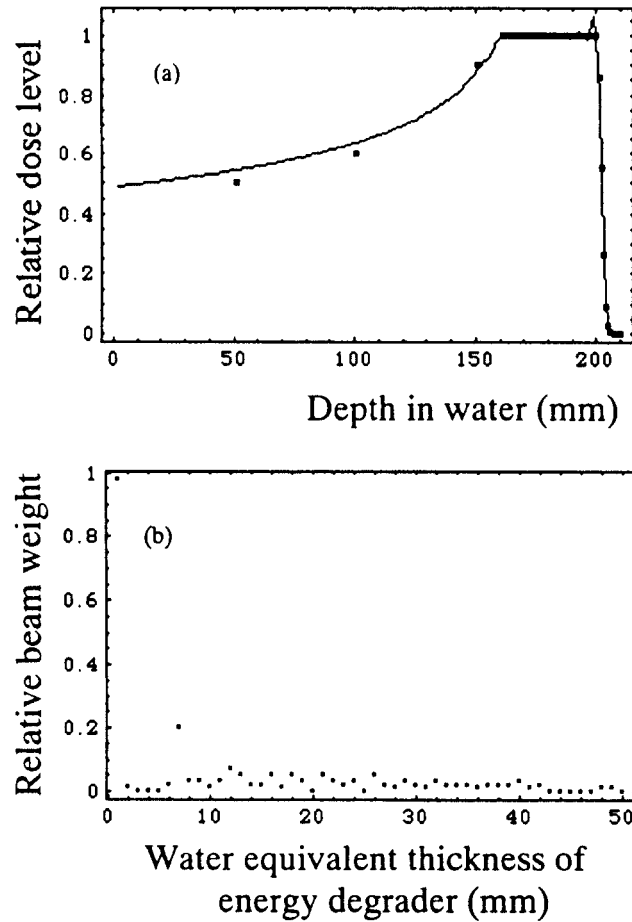
Figure 2. Results of one-dimensional lateral dose optimization: (*a*) the desired beam profile (dots) as compared with the profile obtained by the GSPM procedure (full curve); (*h*) the beam weight pattern required to generate the GSPM dose distribution shown in (*u*); (*c*) the beam profile generated by a uniform beam (thin curve) compared with the profile generated by the GSPM procedure (thick curve).

where x denotes the lateral coordinate, N_i is a beam weight, Δx is a step size for scanning, σ is a $1/e$ beam radius at the depth z_0, and $S(E, z_0)$ is a stopping power at the depth z_0 generated by a broad parallel beam having an initial energy of E.

Figure 2(*a*) shows a desired lateral dose distribution (dots) and the optimized dose distribution (full curve) calculated by the GSPM algorithm, where the desired lateral dose

Figure 3. Results of one-dimensional depth dose optimization. (*a*) The desired beam profile (dots) as compared with the profile obtained by the GSPM procedure (full curve). The number of dose sampling points in the uniform region was 40, and thus it looks like a thick line. (*b*) An optimal beam weight solution as a function of energy degrader thickness, where monoenergetic proton beams of 50 different energies with a constant range decrement of 1 mm irradiate a water sample.

fall-off was given by the Gaussian distribution characteristic for a single pencil beam. Figure 2(*b*) depicts the optimal solution (dots) of lateral beam weights. The full curve was drawn only for easier observation. In figure 2(*c*), the optimized dose distribution (thick curve) is compared with the dose resulting from a uniform beam (thin curve). These results indicate that the optimization minimizes the lateral dose fall-off to the physical limit while providing a homogeneous dose inside the target.

3.2. One-dimensional depth dose optimization

The depth dose distribution on a central axis $D(0, z)$ generated by a broad parallel beam is given as follows:
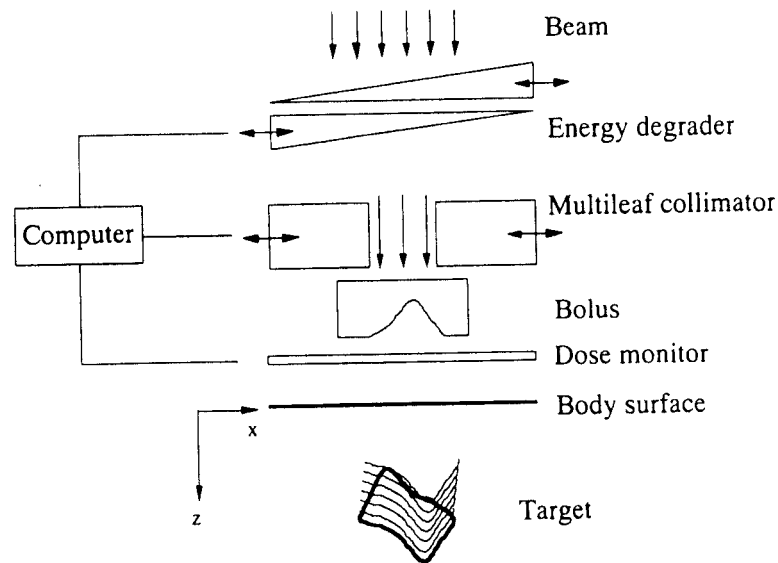
$$D(0, z) = \sum_i N_i S(E_i, z) \tag{4}$$

Figure 4. Diagram of the broad beam three-dimensional proton therapy system.

where z denotes the depth coordinate, N_i is a beam weight of an initial energy E_i, and $S(E_i, z)$ is stopping power at the depth z generated by the beam having the initial energy of E_i. In this paper, the relative stopping power has been taken from measurements made on the 187 MeV broad proton beam from the Gustav Werner synchrocyclotron in Sweden (Larsson 1961).

Figure 3(a) shows a desired depth dose distribution (dots) and the optimized dose distribution (full curve) using the GSPM procedure, where the desired distal dose fall-off was given by the fall-off of a monoenergetic beam having the same range. Monoenergetic proton beams of 50 different energies with a constant range decrement of 1 mm irradiate a water sample. The number of dose sampling points in the uniform region was 40, and thus it looks like a thick full line. Figure 3(b) depicts an optimal beam weight distribution as a function of water equivalent thickness of an energy degrader. The results indicate that the optimization minimizes the distal dose fall-off to the physical limit while giving a homogeneous dose distribution inside the target. The overshoot at the distal corner results from the ideal fall-off characteristic provided as a part of the desired distribution.

As shown in Figure 3(a), we placed only three dose sampling points at the plateau region for the desired dose distribution because we do not know the exact dose level in this region beforehand. The GSPM pattern matching procedure utilizes the vector inner product given by equation (2), and therefore the assigned sampling point density automatically corresponds to the weighting coefficient of the pattern matching calculation.

3.3. Application to broad beam three-dimensional proton therapy

Figure 4 shows a diagram of a broad beam three-dimensional proton therapy system (Kanai et al 1983a). The broad beam that runs in parallel irradiates a target inside a body through a variable energy degrader, a multileaf collimator, a fixed bolus and a dose monitor. A computer receives dose signals from the dose monitor; subsequently, the computer transmits control signals to the energy degrader and the multileaf collimator during irradiation.

In this system, the treatment volume is layered according to the shape of the distal surface of the target, and the dose is built up using the energy degrader, the multileaf
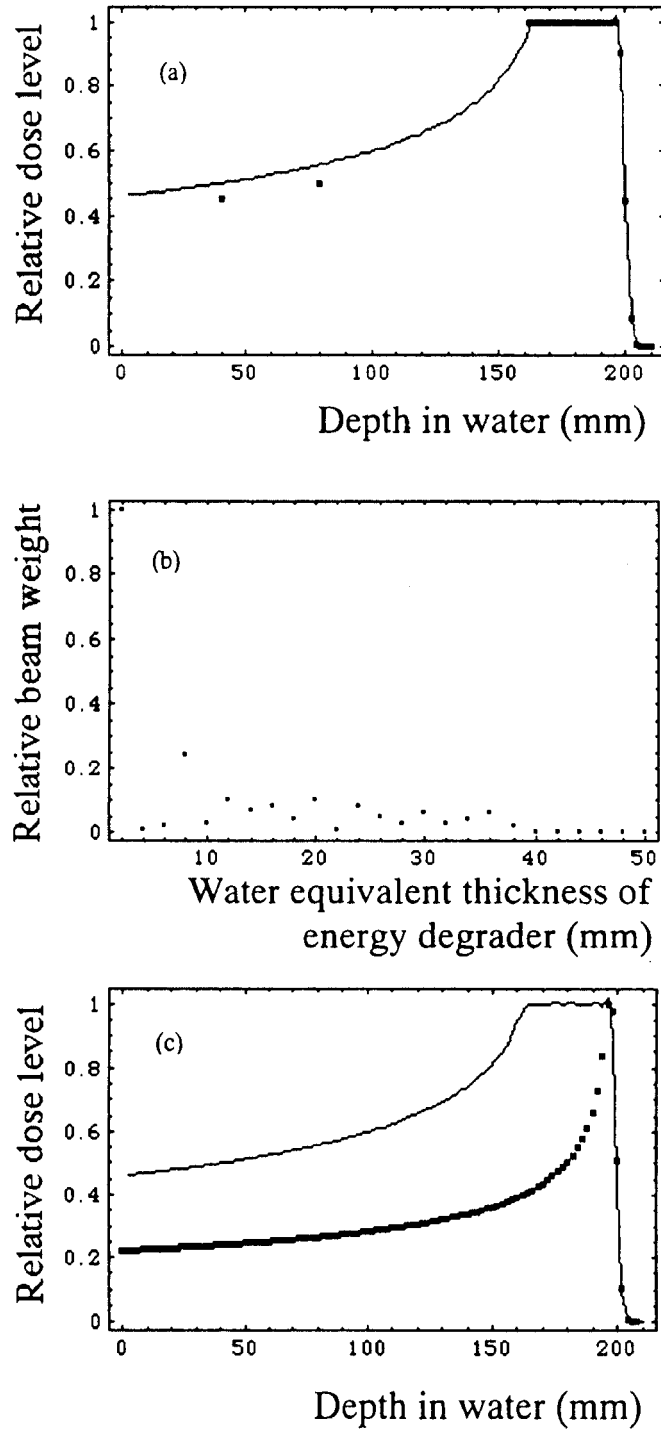
Figure 5. Results of depth dose optimization for the broad beam three-dimensional therapy system. The optimization was performed along the depth-directed line including the deepest point of the target. (*a*) The desired depth dose distribution (dots) is compared with the optimized dose distribution (full curve). The number of dose sampling points in the uniform region was 18, and thus it looks like a thick line. (*b*) An optimal beam weight is shown as a function of energy degrader thickness, where monoenergetic proton beams of 25 different energies with a constant range decrement of 2 mm irradiate a water sample. (*c*) The optimized dose distribution (full curve) is compared with a monoenergetic Bragg curve (dots).

Figure 6. Three-dimensional plots (a)–(d) of the dose distribution during optimized irradiation.

collimator and the bolus. When the range is shortened by increasing the degrader thickness, the irradiated field aperture is changed using the multileaf collimator to maintain conformal irradiation.

The absorbed dose contribution to a point at depth z and distance x from a central axis of a Gaussian proton beam having an energy E and a beam weight $N(E)$ is given as follows:

$$D(x, z) = N(E) \frac{\exp[-x^2/\sigma(z)^2]}{\pi \sigma(z)^2} S(E, z) \tag{5}$$

where $\sigma(z)$ is a $1/e$ beam radius at the depth z and $S(E, z)$ is a stopping power. In this calculation, the initial $1/e$ radius of the incident Gaussian beam is set at 3 mm, and the beam radius in water has been taken from measurements made on the 177 MeV proton beam from the PSI cyclotron in Switzerland (Scheib *et al* 1994). Further, it is known that a broad collimated beam can be generated by a summation over a finite number of narrow Gaussian beams (Brahme *et al* 1981), and this idea was used for the present calculation. The edge scattering at the collimator wall were ignored for simplicity.

Figure 5(a) shows a one-dimensional desired depth dose distribution (dots) and the optimized dose distribution (full curve) using the GSPM procedure, where the optimization was performed along the depth-directed line including the deepest point of the target. The number of dose sampling points in the uniform region was 18, and thus it looks like a thick line. Again the desired distal dose fall-off was given by the fall-off of a monoenergetic beam having the same range. Figure 5(b) depicts an optimal beam weight solution as a function of energy degrader thickness. Monoenergetic proton beams of 25 different energies with constant range decrements of 2 mm irradiate the medium. Figure 5(c) shows the
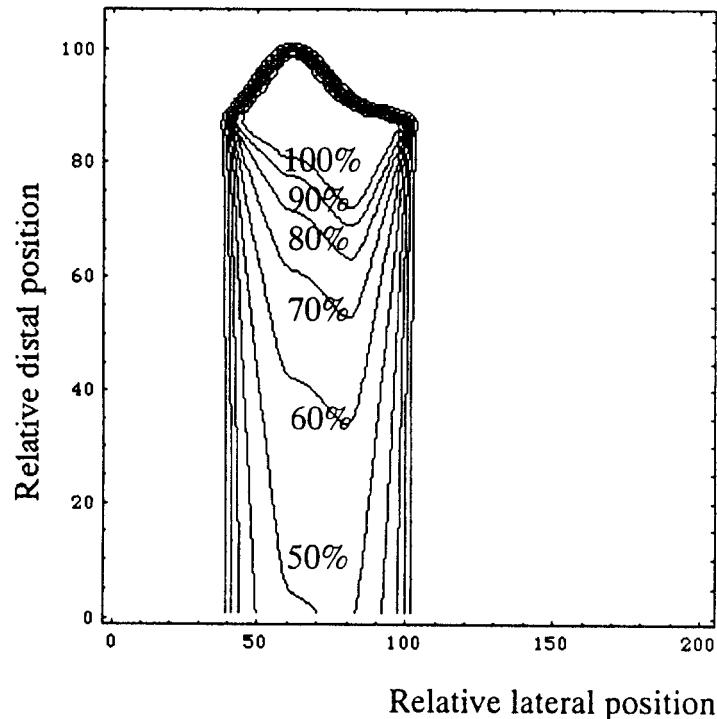
Figure 7. Isocontour plot of the optimized dose distribution.

optimized dose distribution (full curve) and a monoenergetic Bragg curve (dots), indicating that the optimization minimizes the distal dose fall-off while providing a homogeneous dose distribution inside the target.

Figures 6(a)–(d) show three-dimensional plots of the dose distribution during the optimized irradiation, while figure 7 gives an isocontour plot of the final dose distribution. Although we have optimized the dose distribution only in a depth direction, the lateral dose fall-off was also practically optimized due to the target shape characteristic. We suggest that for most convex cancer targets the proposed optimization works well for broad beam three-dimensional proton therapy with an advantage of minimized distal and lateral fall-off. The one-dimensional GSPM calculation time using a Power Macintosh 8100/100 MHz was of the order of 10 s depending on the number of beams considered.

4. Discussion

In this report we show preliminary results using homogeneous water phantom models. However, once we calculate spatially varying energy deposition kernels using Monte Carlo methods or other analytical approximations such as semi-infinite slab models, it is possible to perform dose optimization in a heterogeneous body using the GSPM method. We have demonstrated one-dimensional examples followed by an application to the broad beam therapy system; however, the approach can be directly applied to a general three-dimensional spot scanning method (Kanai *et al* 1980, 1983b). The dose optimization for the three-dimensional spot scanning in heterogeneous media will be reported elsewhere. The GSPM procedure can also be applied to dose optimization of other radiotherapy means including photon, electron and heavier charged ions by considering RBE.

Acknowledgments

The authors are grateful to M Maeda, H Tsuchidate, T Oichi, T Nakanishi, S Nakamura, and K Nakanishi for their support and encouragement. The first author also wishes to thank anonymous referees for many valuable comments and suggestions.

References

Brahme A, Källman P and Lind B 1989 Optimization of proton and heavy ion therapy using an adaptive inversion algorithm *Radiother. Oncol.* **15** 189–97

Brahme A, Lax I and Andreo P 1981 Electron beam dose planning using discrete Gaussian beams *Acta Radiol. Oncol.* **20** 147–58

Goitein M and Sisterson J M 1978 The influence of thick inhomogeneities on charged particle beams *Radiat. Res.* **74** 217–30

Kanai T, Kawachi K, Kumamoto Y, Ogawa H, Yamada T, Matsuzawa H and Inada T 1980 Spot scanning system for proton radiotherapy *Med. Phys.* **7** 365–9

Kanai T, Kawachi K, Matsuzawa H and Inada T 1983a Broad beam three-dimensional irradiation for proton radiotherapy *Med. Phys.* **10** 344–6

——1983b Three-dimensional beam scanning for proton therapy *Nucl. Instrum. Methods* **214** 491–6

Larsson B 1961 Pre-therapeutic physical experiments with high energy protons *Br. J. Radiol.* **34** 143–51

Larsson T 1987 Absorbed dose calculations and dose planning with proton beams using a Gaussian pencil beam *Uppsala University School of Engineering, Department of Technology: Report Series* UPTEC 87116 E

Lind B 1988 Comparison of algorithms for solving inverse problems in radiotherapy *Book of Abstracts VIII ICMP (San Antonio, Texas)* paper MP 5.6

——1990 Properties of an algorithm for solving the inverse problem in radiation therapy *Inverse Probl.* **6** 415–26

Midorikawa Y, Ogawa J, Doi T, Hayano S and Saito Y 1997 Inverse analysis for magnetic field source searching in thin film conductor *IEEE Trans. Magn.* **33** 4008–10

Pedroni E 1995 The 200-MeV proton therapy project at the Paul Scherrer Institute: conceptual design and practical realization *Med. Phys.* **22** 37–53

Russell K, Grusell E and Montelius A 1995 Dose calculation in proton beams: range straggling corrections and energy scaling *Phys. Med. Biol.* **40** 1031–43

Saito Y, Itagaki E and Hayano S 1990 A formulation of the inverse problems in magnetostatic fields and its application to a source position searching of the human eye fields *J. Appl. Phys.* **67** 5830–2

Saito Y and Yoda K 1996 Method for generating energy distribution *Japanese Patent Application* 408315279

Saotome H, Coulomb J-L, Saito Y and Sabonnadiere J-C 1995 Magnetic core shape design by the Sampled Pattern Matching method *IEEE Trans. Magn.* **31** 1976–9

Saotome H, Kitsuta K, Hayano S and Saito Y 1993 A neural behaviour estimation by the generalized correlative analysis *IEEE Trans. Magn.* **29** 1389–94

Scheib S, Pedroni E, Lomax A, Blattman H, Böhringer T, Coray A, Lin S, Munkel G, Schneider U and Tourovsky A 1994 Spot scanning with protons at PSI: experimental results and treatment planning *Hadrontherapy in Oncology* ed U Amaldi and B Larsson (Amsterdam: Elsevier) pp 471–80

Yoda K, Oka T, Tsutaka Y, Usami T, Nakamura H and Umeda Y 1997 Shape estimation of radioactive spent resins in a storage tank using the Sampled Pattern Matching *IEEE Trans. Nucl. Sci.* **44** 1469–73

Wolfram S 1991 *Mathematica, a System for Doing Mathematics by Computer* (San Francisco: Addison Wesley)