INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Signal processing and image restoration techniques for two-dimensional eddy current nondestructive evaluation

by

Bing Wang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering (Communications and Signal Processing)

Major Professor: John P. Basart

Iowa State University

Ames, Iowa

1997

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Bing Wang

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

To my parents,

Shizhong Wang and Xinbi Ge

and my sister,

Hongyin Wang

for their understanding, support, and encouragement

throughout the many years of my school life

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. John P. Basart for his guidance, direction, and support of my research. I have greatly benefited from his knowledge in both electromagnetics and signal processing and his high standard for academic work. This dissertation would not have existed if he had not given me the chance and challenge to work on this hard problem.

I would like to thank John C. Moulder for serving as my supervisor and my committee member. His expertise in the eddy current NDE field has been an invaluable asset to my research.

I would like to thank Dr. James H. Rose for serving as my committee member. I have enjoyed many discussions with him on the nature of physics.

I also like to thank Dr. Donald O. Thompson, Dr. Lalita Udpa, Dr. Julie Dickerson, and Dr. Jennifer Davidson for their help and serving as my committee members.

I am grateful to many students I have worked with, Sheng-Fa Chuang, Zhong Zhang, Lunxiong He, Lavanya Apsani, Han-Chi Hsieh, and Amneh Akour for they have made the research work more enjoyable.

ABSTRACT

This dissertation presents a comprehensive study on the forward modeling methods, signal processing techniques, and image restoration techniques for two-dimensional eddy current nondestructive evaluation. The basic physical forward method adopted in this study is the volume integral method. We have applied this model to the eddy current modeling problem for half space geometry and thin plate geometry. To reduce the computational complexity of the volume integral method, we have developed a wavelet expansion method which utilizes the multiresolution compression capability of the wavelet basis to greatly reduce the amount of computation with small loss in accuracy. To further improve the speed of forward modeling, we have developed a fast eddy current model based on a radial basis function neural network. This dissertation also contains investigations on signal processing techniques to enhance flaw signals in two-dimensional eddy current inspection data. The processing procedures developed in this study include a set of preprocessing techniques, a background removal technique based on principal component analysis, and grayscale morphological operations to detect flaw signals. Another important part of the dissertation concerns image restoration techniques which can remove the blurring in impedance change images due to the diffusive nature of the eddy current testing. We have developed two approximate linear image restoration methods - the Wiener filtering method and the maximum entropy method. Both linear restoration methods are based on an approximate linear forward model formulated by using the Born approximation. To improve the quality of restoration, we have also developed nonlinear image restoration methods based on simulated

annealing and a genetic algorithm. Those nonlinear methods are based on the neural network forward model which is more accurate than the approximate linear forward model.

CHAPTER 1. INTRODUCTION

Overview of the Problem

Eddy current nondestructive evaluation (NDE) is one of the most important NDE techniques. In an eddy current measurement, a coil driven by sinusoidal current is placed on a metal testpiece and the impedance change of the coil is measured as the coil is moved around the testpiece. By observing the change in impedance as a function of coil position, we can detect and characterize defects in the testpiece. Applicable types of defects of the eddy current technique vary from inclusions, voids to surface breaking cracks. Since eddy current testing only requires one side access to the testpiece, it can be used in complex test geometry where many other NDE techniques cannot be easily applied. Typical applications of eddy current testing includes the inspection of steam generator tubing in nuclear power plants and the inspection of surface structures of aircraft. The techniques developed in this dissertation were originally targeted for the tubing inspection application. However, they can be applied to many other two-dimensional eddy current testing applications as well.

In a steam generator in a nuclear power plant there are thousands of heat exchanger tubes. These tubes are used to transfer heat energy from the nuclear reactor to the power generator. These tubes must also separate the primary water which is radioactive from the secondary water which is not radioactive. Therefore, the integrity of these tubes is critical to the safe operation of the nuclear power plant. Because of the relatively small size of the heat

exchanger tubes and the accessibility constraints of the steam generator, eddy current testing is widely used in the inspection of the heat exchanger tubes.

Traditionally a bobbin coil probe is used to inspect steam generator tubing. During the inspection the bobbin coil is inserted inside the tube and then pulled out. While the probe is moving along the internal surface of the tube, the impedance change of the coil is monitored and recorded. From the one-dimensional impedance change data obtained from the bobbin coil, we can decide whether there are flaws and approximately how large the flaws are. Most flaws in steam generator tubing happen to be circumferential outside diameter (OD) cracks near tube support plates. Because of the one-dimensional nature of the bobbin coil, it has a limited sensitivity to circumferential cracks. Therefore, the use of Motorized Rotating Pancake Coils (MRPC) for eddy current inspection has become a more common practice in recent years, as utilities struggle to stay ahead of cracking steam generator tubes. In a MRPC inspection, a small pancake coil is placed on the internal surface of the tube and rotated while the MRPC probe is pulled out. Therefore, the trace of the pancake coil has a helical shape. MRPC probes can provide two-dimensional impedance change images over the tube internal surface which improve the probability of detection and enhance flaw characterization capability.

Despite its performance improvement over the bobbin coil, the MRPC still has some disadvantages. One of its major disadvantages is sensitivity-limiting noise resulting from stray electromagnetic pickup and irregular probe motion. Another major disadvantage is the blurring of the impedance change image due the nonlinear interaction between the flaw and the probe coil. Therefore, to improve the quality of data analysis, signal and image

processing techniques must be developed to reduce noise and unwanted effects and to restore the actual flaw shape from the blurred impedance change image.

The major objective of the study presented in this dissertation is to develop signal and image processing techniques for the two-dimensional eddy current data so that better flaw detection and characterization can be achieved. This includes signal processing techniques which enhance flaw signals and image restoration techniques which generate more accurate flaw images. Also, to have a better understanding of the problem, we must study the physical phenomenon involved in eddy current testing. This requires us to solve the eddy current forward modeling problem.

Fundamentals of Eddy Current Testing

Basic concept

Fig. 1.1 illustrates the basic concept of eddy current nondestructive testing. A varying electric current flowing in a coil gives rise to a varying magnetic field. A nearby conductor resists the effect of the varying magnetic field, and this manifests itself by an eddy current flowing in a closed loop in the surface layer of the conductor to oppose the change causing a back electromotive force in the coil. Cracks and other surface conditions modify the eddy currents generated in the conductor so that the back electromotive force is altered. This leads to a change in the impedance of the coil. If the conductor is a metal testpiece with defects, we can detect these defects by moving the coil around the surface of the testpiece and measuring the impedance change of the coil.

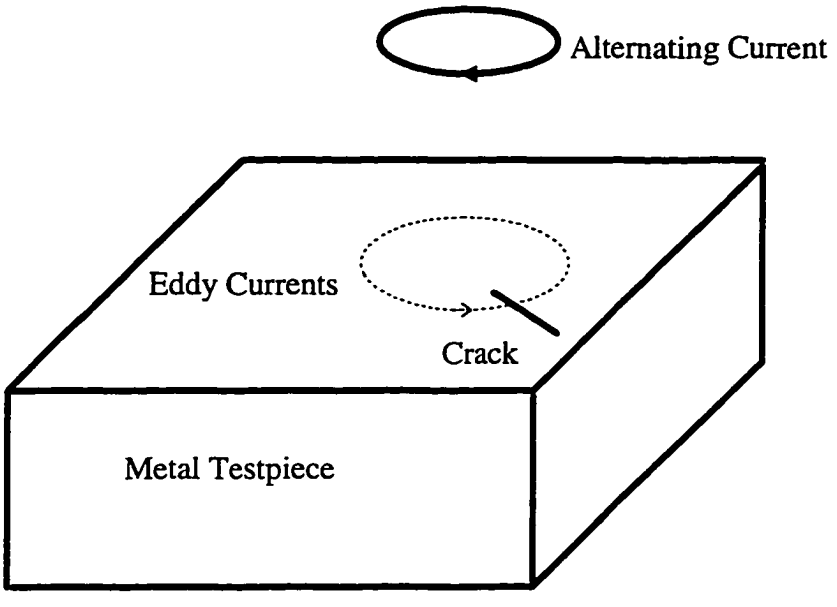


Fig. 1.1. Basic concept of eddy current nondestructive testing.

Practical considerations

Penetration depth

Since the eddy current electromagnetic fields are diffusive fields, they have limited penetration into the metal testpiece. The rate of decay of the eddy currents is measured by the skin depth, which is dependent on the test frequency, the permeability of the testpiece, and the conductivity of the testpiece. Since the permeability and conductivity are material properties, in eddy current inspection we can only change the test frequency to control the penetration depth of eddy current testing. Because flaws deeper than two to three times of the skin depth usually lead to very small responses in the coil impedance change, we need to make the test frequency low enough so that all structures of interests are sensible by the coil.

However, because the resolution of eddy current testing is also dependent on the skin depth, we can only achieve high resolution of near surface structures.

Multiple frequency inspection

Due to the frequency dependent nature of eddy current testing, we can use multiple frequency channels to improve its flaw detection capability. Because high frequency channels are more sensitive to near surface structures and low frequency channels are more sensitive to structures deeper in the testpiece, we can combine their results to improve our capability of differentiating types of flaw. Furthermore, multiple frequency inspections enable the mixing process, which can be used to remove signals from permanent structures of the testpiece; so that, signals from flaws adjacent to these permanent structures are more distinguishable.

Complex data display

Due to the impedance change of the coil being a complex quantity, there is a phase component. The phase component of the impedance change is very important in eddy current testing because it is a function of flaw depth. By measuring the phase of the acquired data, we can differentiate near surface flaws and deep flaws. It is also possible to estimate the depth of a surface breaking flaw from the phase of its signal. In eddy current data analysis, the complex impedance change data is usually plotted as a Lissajous figure, with the horizontal axis being the real part of the impedance change and the vertical axis being the imaginary part of the impedance change. By observing the shape and the angle of the

Lissajous pattern of a signal, a human analyst can decide whether there is a flaw and approximately how large the flaw is.

Tubing inspection

Eddy current tubing inspection is different from the inspection of a flat surface testpiece because there are more uncertainties in the test. First, there could be changes in the diameter of the tube and vibrations of the probe when it is pulled out from the tube, causing large liftoff signals in the tube inspection data. Second, tube support plates (TSPs) in the steam generator create large signals that may obscure signals from flaws near the support plates. Third, many other changes of the tube over time, e.g., dents, deposits, and wear, can create signals that are similar to crack signals. This makes the analysis of the tubing inspection data even more difficult.

In a tube, a flaw can be either on the internal surface of the tube, which is called an inside diameter (ID) flaw, or on the outside surface of tube, which is called an outside diameter (OD) flaw. The depth of the flaw is usually measured in percentages of tube through wall (TW) thickness. Most flaws in steam generator tubing happen to be OD circumferential cracks near tube support plates.

Organization of the Dissertation

Corresponding to the three major tasks of this study, the dissertation is divided into three major parts. In Chapters 3-6 we discuss eddy current forward modeling techniques for half space and thin plate geometry. The techniques discussed include an implementation of

the volume integral method for half space and thin plate geometry, a wavelet expansion method to reduce the computational complexity of the volume integral method, and artificial neural network based forward models to substantially speed up the forward computation. In Chapters 7-9 we discuss signal and image processing techniques to enhance flaw signals and to detect possible flaw indications. The processing techniques discussed in this part include preprocessing methods to remove the liftoff signal and to adjust the phase of the signal, a method based on principal component analysis to remove background signals in the data, and a detection method based on grayscale morphological operations to find circumferential cracks. In Chapters 10-13 we discuss eddy current image restoration techniques. The image restoration techniques discussed in this part include linear restoration techniques based on Wiener filtering and the maximum entropy method, and nonlinear image processing techniques based on simulated annealing and genetic algorithms.

In Chapter 2 we review some of the modeling methods and signal processing techniques in the literature that are related to the processing and analysis of two-dimensional eddy current data. The review is divided into three major parts: (1) analytical and numerical methods to model the eddy current flaw-coil interaction and to predict coil impedance change; (2) signal processing techniques to enhance flaw signals; and (3) solutions of the eddy current inverse problem to obtain flaw characteristics from impedance change measurements.

In Chapter 3 we briefly review the mathematical formulation of the volume integral method. We focus the discussion on how to apply the volume integral method to half space geometry and thin plate geometry because these two geometry are sufficient to model most

eddy current applications, either precisely or approximately. The analytical solutions for the incident fields and the Green's functions for half space geometry and thin plate geometry are also discussed.

In Chapter 4 we discuss a practical implementation of the volume integral method for half space geometry and thin plate geometry. The implementation of the volume integral method is difficult because it involves calculations of complex and numerically unstable integrations of Bessel functions. We applied fast Hankel transforms to efficiently integrate many of the integrations used in the volume integral method. In this chapter we also discuss how to compute the volume integral of Green's function in the singular element which is critical to the accuracy of the forward model.

In Chapter 5 we discuss a wavelet expansion method to reduce the computational complexity of the volume integral method. This method utilizes the multiresolution compression capability of the wavelet basis and the diffusive nature of eddy currents. By using a wavelet basis in a Galerkin method, we transform the system matrix to a very sparse matrix which can be inverted efficiently. The wavelet expansion method has been shown to greatly reduce the amount of computation involved in the volume integral method with only a small loss in accuracy.

In Chapter 6 we discuss fast eddy current models based on artificial neural networks. Since numerical models are inherently computationally intensive, they cannot be used in situations where speed is the first concern and accuracy is a secondary concern. To improve the speed of forward modeling, we utilized the functional approximation capability of neural networks to create an approximate mapping from the flaw image to the impedance change

image from a set of training samples. The neural network based forward models have shown to be able to provide speedup of several orders as compared to numerical models.

In Chapter 7 we discuss preprocessing techniques to cleanup the two-dimensional eddy current data. The preprocessing techniques include background removal methods based on polynomial fitting and median filtering, and an automatic phase adjustment method based on a least squares criterion.

In Chapter 8, a background removal method based on principal component analysis (PCA) is discussed. In this technique, PCA is used to extract major components of the background signals in the data. Flaw signals are considered as details of the image and can be separated from the background by using the result of the PCA processing. This technique has been shown to be effective in removing noise from ID variations and manufacturing induced conditions.

In Chapter 9 we discuss a detection method for circumferential cracks based on grayscale morphological operations. This method is based on two shape characteristics of the signal from a circumferential crack: minimum horizontal extent and maximum vertical extent. These shape characteristics can be recognized by mathematically morphological operations. By using a serial combination of different morphological operations, circumferential cracks in the MRPC data can be enhanced and a crack map can be generated to indicate possible flaw locations.

In Chapter 10, an eddy current image restoration technique based on Wiener filtering and an approximate linear eddy current forward model are discussed. The linear forward model is based on a reciprocity impedance change formula and the Born approximation. This

method has been tested using both synthesized and experimental data. Its sensitivity to noise, its strengths, and limitations are also discussed in Chapter 10.

In Chapter 11, another linear eddy current image restoration technique - the maximum entropy method (MEM) is discussed. MEM has been successfully applied in many areas of image processing. The principle of maximizing the entropy has been shown to provide exceptional results in many applications. In this chapter, the concepts and the mathematical formulation of MEM for the eddy current image restoration problem are discussed. Test results on synthesized data, experimental data, and inspection data have all shown its superior performance over the Wiener filtering method in terms of both noise reduction and resolution. The major disadvantage of the MEM approach is its high computational complexity when compared with the Wiener filtering approach.

To improve the quality of the restoration results, we have also developed nonlinear image restoration techniques which are based on the fast neural network forward model. In Chapter 12 we discuss a nonlinear image restoration technique based on simulated annealing. Although this method is more computationally intensive than the linear restoration methods, it can provide restoration results with better resolution and a higher signal-to-noise ratio.

In Chapter 13 we discuss another nonlinear image restoration methods based on a genetic algorithm. Although the procedure of the genetic algorithm based approach is more complicated than the simulated annealing based approach, it can generate similar restoration results with a much faster convergence speed due to the parallel search property and the crossover operation of the genetic algorithm. This method provides a better tradeoff between the quality of the restoration and the execution time.

The dissertation is summarized and concluded in Chapter 14. Major results of the study are reviewed and the strengths and weaknesses of each method are discussed and compared. In this chapter we also discuss what we have learned during the course of the study and some suggestions for future research.

CHAPTER 2. LITERATURE REVIEW

Introduction

In this chapter we briefly review modeling methods and processing techniques in the literature that are related to the processing and analysis of two-dimensional eddy current data. The review is divided into three major parts: (1) analytical and numerical methods to model the eddy current flaw-coil interaction and to predict coil impedance change; (2) signal processing techniques to enhance flaw signals; and (3) solutions of the eddy current inverse problem to obtain flaw characteristics from impedance change measurements. The first part of the review involves solutions of the eddy current forward problem which provide an understanding of the physical criteria associated with eddy current testing. The second part of the review involves signal enhancing techniques to improve the quality of experimental or inspection data; thus, flaws can be more easily located and better characterized. The third part of the review involves solutions of the eddy current inverse problem which are directly related to the flaw characterization capability of the eddy current nondestructive testing technique. The three parts of the problem are closely related: although the final results are generated by methods that solve the inverse problem, our understanding of the first two parts are critical to how well we can solve the inverse problem. Our objective of maximizing the inspection capability of the eddy current technique can only be achieved after we fully understand all the aspects associated with this problem.

Forward Modeling

Overview

The eddy current forward modeling problem has been an active research area for a relatively long time. Good understanding and accurate modeling of the forward problem have a significant impact on the design of new probes and inspection instruments as well as the solution of the inverse problem.

The eddy current forward problem can be considered as the problem to solve the incident electromagnetic field distribution, which is the field when there is no flaw in the test object, and the total electromagnetic field distribution, which is the field when there is a flaw in the test object. This is based on Auld's well known reciprocity formula [1] for the impedance change of a coil placed above a metal testpiece

$$\Delta Z = -\frac{1}{I^2}\int_{V'} \delta\sigma(\mathbf{r}')\mathbf{E}^0(\mathbf{r}')\cdot\mathbf{E}(\mathbf{r}')dv', \tag{2.1}$$

where $\mathbf{E}^0$ is the incident electric field, $\mathbf{E}$ is the total electric field, I is the current density in the coil, $\delta\sigma(\mathbf{r}')$ is the conductivity change, and V' is the flaw volume.

At the frequency of typical eddy current inspection, the electromagnetic field in the metal testpiece is governed by a diffusion equation. If we choose the electric field as the unknown quantity, the diffusion equation can be expressed as [2]

$$\nabla^2\mathbf{E} - \gamma^2\mathbf{E} = i\omega\mu\mathbf{J}^0. \tag{2.2}$$

where $\gamma^2 = i\omega\mu\sigma$, and $\mathbf{J}^0$ is the current density induced by the coil without the presence of the flaw. When the test geometry is simple and the flaw has a regular shape, it is possible to find an analytical solution for the total electric field. However, in most cases numerical

methods must be used to solve for the total electromagnetic field. There are a number of numerical methods in the eddy current literature, e.g., the boundary element method [3, 4], the volume integral method [2, 5, 6], and the finite element method [7]. Although these numerical methods give exact solution of the field, they are inherently computationally intensive. Therefore, in the literature there are also various approximate methods that are valid under certain flaw and test geometry configurations [8-13]. Besides the above mentioned methods which are solidly based on electromagnetic field theory, there is an imaging model by Groshong [14] which is based on a resistive loop approximation and the idea of layered nonlinear representation.

The volume integral method

The volume integral method [2, 5, 6] transforms the governing diffusion equation for eddy current into a volume integral equation which is more suitable for numerical solution. Consider a special solution of (2.2) when the current source is a point electric dipole placed in the layer containing the flaw

$$\nabla^2 \overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') - i\omega\mu\sigma_0 \overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = i\omega\mu\delta(\mathbf{r}-\mathbf{r}')\mathbf{I}, \tag{2.3}$$

where the solution of the above equation $\mathbf{G}(\mathbf{r},\mathbf{r}')$ is the Green's function, $\delta(\mathbf{r}-\mathbf{r}')$ is the three-dimensional Dirac delta function, and $\mathbf{I}$ is the unit dyad. Given the Green's function for the test geometry, the solution of (2.2) can be expressed as a volume integral in the flaw region. To see this, rewrite (2.2) as

$$\nabla^2 \mathbf{E} - i\omega\mu\sigma_0 \mathbf{E} = i\omega\mu[\mathbf{J}^0 + (\sigma-\sigma_0)\mathbf{E}] \tag{2.4}$$

The term $(\sigma - \sigma_0)\mathbf{E}$ can be considered as an effective current dipole density at the source point due to the variation of the flaw conductivity from the host conductivity. We can combine (2.3) and (2.4), and express the total field as

$$\begin{aligned} \mathbf{E}(\mathbf{r}) &= \int \bar{\mathbf{G}}(\mathbf{r},\mathbf{r}') \cdot [\mathbf{J}^0(\mathbf{r}') + \delta\sigma(\mathbf{r}')\mathbf{E}(\mathbf{r}')] dv' \\ &= \mathbf{E}^0(\mathbf{r}) + \int_{V'} \delta\sigma(\mathbf{r}')\bar{\mathbf{G}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}') dv' , \end{aligned} \tag{2.5}$$

Given the Green's function and the incident field $\mathbf{E}^0$, we can solve the above volume integral equation for the total field $\mathbf{E}$. This is usually done by discretizing equation (2.5).

Compared with other numerical methods, the volume integral method is easy to implement and it can model three-dimensional flaws with arbitrary conductivity distribution. The disadvantage of the volume integral method is that it uses a large number of volume elements, and thus has high computational resource requirement.

The boundary element method

The boundary element method [3, 4] is formulated by an application of Green's theorem to the unknown field and to a Green's function to obtain the solution. If the Laplace Green's function is used in solving a Helmholtz equation, the field at a point is expressed as a volume integral over the region of interest. This leads to the volume integral method. However, if a Helmholtz Green's function is used, the field at a point is expressed by integrals over a bounding surface. If we let the field point move to the surface, we obtain an integral equation over the surface for the unknown fields, where the kernels are the Helmholtz Green's functions and their normal derivatives. If we let $\mathbf{H}$ be the unknown field, then the boundary integral equation can be derived as [3]

$$\Delta \mathbf{H}(\mathbf{r}) = \int_{S'} \overline{\mathbf{G}}(\mathbf{r},\mathbf{S}')\Delta \frac{\partial \mathbf{H}(\mathbf{S}')}{\partial \mathbf{n}'} d\mathbf{S}', \tag{2.6}$$

where S' is the flaw region, and $\overline{\mathbf{G}}(\mathbf{r},\mathbf{S}')$ is the Helmholtz Green's function for the given geometry. Equation (2.6) can be discretized and solved numerically for the unknown **H** field, which can then be used in computing the impedance change.

The major advantage of the boundary element method is that it reduces a three-dimensional problem to a two-dimensional problem so that a lot of computation can be saved. However, the Helmholtz Green's function is more complex than the Laplace Green's function. For some test geometry, it may not be possible to solve the Helmholtz Green's function analytically. It is also more difficult for the boundary element method to model a complete three-dimensional problem than the volume integral method.

The finite element method

The finite element method is a general method of solving partial differential equations (PDEs). The basic idea of the finite element method is to use variational principles to transform a boundary value problem into a set of linear equations. A widely used method is the Galerkin method [7]. In the Galerkin method, the unknown field is expressed as a series expansion based on a set of basis functions. The basis functions are selected to satisfy the boundary conditions. By substituting the series expression of the unknown field into the partial differential equation and utilizing the orthogonality property of the basis functions, we can transform the boundary value problem into a set of linear equations which can be solved numerically.

The advantages of the finite element method are its high accuracy and its capability of modeling arbitrary test geometry and flaw configuration. The disadvantages of the finite element method are its high complexity and high computational resource requirement.

The Groshong model

The Groshong model is an eddy current imaging model based on a simple resistive loop approximation. This model maps a two-dimensional flaw image to a two-dimensional magnitude image of the impedance change of the coil. In this model, the interaction between the coil and the flaw is considered as a transformer where the coil is the primary of the transformer and the current filaments induced by the coil are considered the secondary of the transformer. A flaw in the metal testpiece changes the length of circuit path of the current filament, thus changing the resistance of the secondary and creating a voltage change in the primary. The mathematical formulation of the Groshong model uses the layered nonlinear representation. The first layer of the model is a directional first derivative of a Gaussian function

$$H(x,y,\theta) = -\frac{(x\cos\theta + y\sin\theta)}{\sqrt{2\pi}\sigma_d}\exp\left\{-\frac{(x\cos\theta + y\sin\theta)^2}{2\sigma_d^2}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_g}\exp\left\{-\frac{(-x\sin\theta + y\cos\theta)^2}{2\sigma_g^2}\right\} \tag{2.7}$$

in the rotated two-dimensional Cartesian coordinate system $(x', y') = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta)$ at the angle θ. The first term of the right hand side is the first derivative of a Gaussian function with spread σ_d parallel to the direction defined by

θ, and the second is a Gaussian function with a much larger spread σ_g perpendicular to θ. This function is used to reflect the fact that the coil impedance change is more sensitive to the flaw perpendicular to the current filaments, and less sensitive to the flaw parallel to the current filament. A number of H functions with different angles θ, are used to approximate the coil response in the full angle range.

The second layer of the model is a nonliear point function

$$s(v) = \sqrt{1+v^2} \tag{2.8}$$

which is used to estimate the local path length of filaments flowing in the direction of θ.

The third layer is a linear Gaussion blurring function

$$R(r,\phi) = \frac{1}{2\pi\sigma_r\sigma_\phi}\exp(-\frac{(r-r_0)^2}{2\sigma_r^2} - \frac{(\phi-\phi_0)^2}{2\sigma_\phi^2}), \tag{2.9}$$

where σ_r is the standard deviation of the blur in the radial direction, and σ_ϕ is the angular extent. This function is used to smooth out the effect of using a finite number of H functions to approximate the response in the full angle range.

The Groshong model is relatively simple when compared to models based on electromagnetic theory and it is less computationally intensive. Also, because it uses a layered nonlinear form, we can easily obtain its derivative. Thus a gradient based inverse method can be used for this forward model. However, the Groshong model also has several disadvantages. First, there is no direct relationship between the coil parameters and the model parameters. Therefore, the model parameters can only be obtained by obtaining a lot of measurements, and then using a nonlinear optimization method to determine the estimates of model parameters. Second, the coil types and test geometry that can be modeled by this

method are limited. Third, because it is a real model, phase information is not used in the model. This is undesirable because in many applications phase information is critical.

Signal Processing Techniques

Overview

The major task in signal processing of eddy current data is to remove noise from various sources, and unwanted signals from non-ideal test conditions and structural changes of the test piece other than the flaw of interest. Due to the complexity of the nonliear interaction between the flaw signal and the unwanted signals, signal processing of eddy current data is a very challenging task. Many conventional signal processing techniques are not efficient in handling eddy current data Traditional eddy current signal processing is mainly based on the phase difference between the flaw signal and unwanted signals. For instance, the signal from an OD crack has a different phase from the liftoff signal. A process called mixing [15] is widely used in industry to separate the flaw signal from signals of unwanted effects. The mixing process utilizes the frequency dependent response of eddy currents to differentiate changes in the test piece. Recently, a spline smoothing technique [16] has been applied to remove random noise in experimental eddy current data. This method is rather effective in reducing electronic noise in experimental data.

Mixing

Mixing is based on the observation that the response of a eddy current probe to a defect in the test piece is frequency dependent. If we measure the same test piece using two

different frequencies, we can combine the signals from the two measurements and enhance the flaw signal. Suppose the flaw signal is larger at the higher frequency, and the unwanted signal is larger at the lower frequency. Let $\mathbf{X}_1(n)$ be the total signal from the higher frequency, and $\mathbf{X}_2(n)$ be the total signal from the lower frequency, thus we can enhance the flaw signal by generating a mixing signal

$$\mathbf{X}_3(n) = \mathbf{X}_1(n) - \mathbf{C}\mathbf{X}_2(n), \tag{2.10}$$

where the complex mixing constant, $\mathbf{C}$, can be decided by minimizing the mean square error

$$\varepsilon = \sum_n \left|\mathbf{X}_1(n) - \mathbf{C}\mathbf{X}_2(n)\right|^2. \tag{2.11}$$

This minimization problem can be easily solved by letting the partial derivative of equation (2.11) with respect to $\mathbf{C}$ be zero and then solving for $\mathbf{C}$. The result is

$$\mathbf{C} = \frac{\sum_n \mathbf{X}_1(n)\mathbf{X}_2^*(n)}{\sum_n \left|\mathbf{X}_2(n)\right|^2}. \tag{2.12}$$

The above mixing algorithm is called a two-channel linear mixing algorithm. There are also three-channel mixing algorithms and nonlinear mixing algorithms. The mixing process is widely used in removing signals from support structures in tubing inspection. However, the mixing process is not very effective in separating the flaw signal and signals from structures with similar material property and depth as the flaw.

Spline smoothing

The spline smoothing method discussed in Reference 16 is an extension to the widely used spline smoothing process. Smoothing splines can be used to approximate a set of

discrete, noisy data. The optimal spline can be obtained by finding the spline coefficients that minimizes a cost function. Usually the cost function contains the average square error between the noisy data and the spline approximation, and a regularization term. The method presented in Reference 16 provides a practical, effective way for estimating the optimal amount of smoothing from the data. Applications of this method have been shown to be very effective in removing the electronic noise in experimental measurements.

The Inverse Problem

Overview

Contrary to the relative success of the forward problem, there have been few practical solutions to the eddy current inverse problem. The reason comes from the physical nature of the eddy current testing; the eddy current inverse problem is an ill-posed nonlinear inverse problem. Due to the diffusive nature of the electromagnetic fields in the frequency range of eddy current measurements, it is difficult to obtain a high-resolution flaw image from a high-resolution measurement.

There are a number of flaw inversion methods in the eddy current literature. Basically they can be classified as parametric methods [17-22], and nonparametric methods [23-27]. In the first class, the flaw shape or the testpiece shape is assumed to be known, and the inverse problem is modeled by several size parameters and material parameters. Usually a table is built from multiple solutions of the forward problem, either experimentally or theoretically, and the inversion is basically to look up the table and find the best match. The table merely represents a mapping from a flaw parameter space to an impedance change feature space.

Therefore, it may have many forms from a simple look-up table [17] to a fuzzy classifier or a neural network [18, 19].

In the nonparametric method class, there are methods [23, 24] based on the Born approximation which simplify the nonlinear inverse problem to a linear inverse problem. However, this approximation is valid only when the conductivity of the flaw is close to the host conductivity or when the aspect ratio of the flaw is within a certain range. More recently, attention has been turned to the nonlinear inverse problem. Groshong et al. [26] have developed an eddy current image restoration method based on constrained gradient descent and the layered nonlinear forward model they developed. Norton and Bowler [27] have derived the expression of the gradient of impedance change to the changes in the flaw conductivity distribution. This gradient expression is based on fundamental electromagnetic theory and requires the solution of a normal forward problem and the solution of a hypothetical forward problem. Based on this gradient expression, a class of gradient-based iterative inverse methods can be used to reconstruct the flaw shape or flaw conductivity distribution. However, because the eddy current inverse problem is highly ill-conditioned, gradient-based methods are very likely to stick to local minima even if regularization constraints are used.

Constrained gradient descent based on the Groshong model

This inversion method is an image restoration technique based on the Groshong forward model. The objective of the restoration is to determine the estimate of the flaw image which minimizes the squared error

$$\|\mathbf{e}\|^2 = \sum_{i,j}\left\{(g(i,j) - D(i,j)(\hat{\mathbf{f}}))^2 + \lambda(C(\hat{f}(i,j)))^2\right\}, \tag{2.13}$$

where $g(i,j)$ is the impedance change measurement, $\hat{f}(i,j)$ is the estimate of the flaw image, $D(i,j)$ is the impedance change image given by the forward model, $C(\hat{f}(i,j))$ is the constraint function, and λ is the Lagrange multiplier. The constraint function can have various forms based on the prior knowledge about the flaw. For example, the constraint function can be defined to minimize the volume of the flaw or to make the flaw as smooth as possible. The minimization of this squared error can be solved iteratively by using a gradient descent method

$$\hat{\mathbf{f}}(m+1) = \hat{\mathbf{f}}(m) - \alpha\frac{\partial\|\mathbf{e}\|^2}{\partial\hat{\mathbf{f}}}, \tag{2.14}$$

where α is the step size. It is straightforward to apply the gradient descent method to the Groshong model because the gradient of the layered nonlinear representation can be easily obtained by using the chain rule.

Besides the constraint term in the squared error, other regularization techniques may also be used to force the gradient descent method, which seeks a local minima of the squared error, into regions containing better solutions. One such technique is the so-called spectral magnitude clipping in which a masking function is defined to represent a desired limiting shape of the spectrum of the estimated flaw image. During each iteration, the spectrum of the updated flaw image is calculated and the spectrum is clipped according to the masking function. This technique is useful in removing the high frequency noise generated in the inversion process.

Although the image restoration method discussed above is relatively simple and not very computationally intensive, it is still limited by the disadvantages of its forward model and the fact that the gradient-based method is unlikely to be able to find the global minimum. Application of this method to real-world problems requires great effort in finding the optimal parameters for the forward model. Modeling errors introduced by inaccurate parameters may greatly deteriorate the restoration result.

Gradient descent methods by Norton and Bowler

Starting from basic electromagnetic theory of eddy current, Norton and Bowler [27] have derived an expression to compute the gradient of the impedance change to the flaw conductivity distribution. This gradient expression requires solving the total electric field for the normal forward problem and then solving the total electric field for another hypothetical forward problem with an adjoint Green's function [27], e.g.,

$$\nabla Z(\mathbf{r}) = -\sigma_0 \mathbf{E}(\mathbf{r}) \cdot \tilde{\mathbf{E}}(\mathbf{r}), \tag{2.15}$$

where $\mathbf{E}(\mathbf{r})$ is the total electric field for the normal forward problem, and $\tilde{\mathbf{E}}(\mathbf{r})$ is the total electric field for the hypothetical forward problem with the adjoint Green's function. In the following we give a brief review of the key derivation of Norton and Bowler's result.

For a unit current source, the reciprocity formula gives the impedance change of the coil as

$$\begin{aligned} Z &= -\int_V \delta\sigma(\mathbf{r}')\mathbf{E}^0(\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}')d\mathbf{r}' \\ &= -\int_V \mathbf{E}^0(\mathbf{r}') \cdot \mathbf{P}(\mathbf{r}')d\mathbf{r}', \end{aligned} \tag{2.16}$$

where $\mathbf{P}(\mathbf{r}') = \sigma_0 \nu(\mathbf{r}')\mathbf{E}(\mathbf{r}')$, and $\nu(\mathbf{r}')$ is the normalized conductivity change. Then it follows

$$dZ = -\int_V \mathbf{E}^0(\mathbf{r}') \cdot d\mathbf{P}(\mathbf{r}')d\mathbf{r}'. \tag{2.17}$$

And from the volume integral equation (2.5), we have a similar equation for the equivalent current density $\mathbf{P}(\mathbf{r})$

$$\mathbf{P}(\mathbf{r}) = \mathbf{P}^0(\mathbf{r}) + \nu(\mathbf{r})k^2 \int_V \mathbf{G}(\mathbf{r},\mathbf{r}') \cdot \mathbf{P}(\mathbf{r}')d\mathbf{r}', \tag{2.18}$$

where $k^2 = i\omega\mu_0\sigma_0$. Taking the derivative of (2.18), we have

$$d\mathbf{P}(\mathbf{r}) = \delta\nu(\mathbf{r})\sigma_0\mathbf{E}(\mathbf{r}) + \nu(\mathbf{r})k^2 \int_V \mathbf{G}(\mathbf{r},\mathbf{r}') \cdot d\mathbf{P}(\mathbf{r}')d\mathbf{r}'. \tag{2.19}$$

Now consider a hypothetical forward problem defined by

$$\mathbf{E}^0(\mathbf{r}') = \tilde{\mathbf{E}}(\mathbf{r}') - k^2 \int \tilde{\mathbf{G}}(\mathbf{r}',\mathbf{r}) \cdot \tilde{\mathbf{E}}(\mathbf{r})\nu(\mathbf{r})d\mathbf{r}, \tag{2.20}$$

where $\tilde{\mathbf{E}}(\mathbf{r}')$ is the total electric field for the hypothetical forward problem, and $\tilde{\mathbf{G}}$ is the adjoint of $\mathbf{G}$; that is

$$\tilde{\mathbf{G}}(\mathbf{r}',\mathbf{r}) = \mathbf{G}^T(\mathbf{r},\mathbf{r}'), \tag{2.21}$$

where ^{T} denotes the transpose of the dyad. For this hypothetical problem, we have a similar equation as (2.18)

$$\tilde{\mathbf{P}}(\mathbf{r}) = \mathbf{P}^0(\mathbf{r}) + \nu(\mathbf{r})k^2 \int_V \tilde{\mathbf{G}}(\mathbf{r},\mathbf{r}') \cdot \tilde{\mathbf{P}}(\mathbf{r}')d\mathbf{r}'. \tag{2.22}$$

Now substituting (2.20) into (2.17), interchanging orders of integration and using (2.21) gives

$$\begin{aligned} dZ &= -\int \tilde{\mathbf{E}}(\mathbf{r}') \cdot d\mathbf{P}(\mathbf{r}')d\mathbf{r}' + k^2 \int \nu(\mathbf{r})\tilde{\mathbf{E}}(\mathbf{r}) \cdot \left[\int \mathbf{G}(\mathbf{r},\mathbf{r}') \cdot d\mathbf{P}(\mathbf{r})d\mathbf{r}\right]d\mathbf{r} \\ &= -\int \tilde{\mathbf{E}}(\mathbf{r}) \cdot \left[d\mathbf{P}(\mathbf{r}) - \nu(\mathbf{r})k^2 \int \mathbf{G}(\mathbf{r},\mathbf{r}') \cdot d\mathbf{P}(\mathbf{r}')d\mathbf{r}'\right]d\mathbf{r}. \end{aligned} \tag{2.23}$$

The quantity in brackets is seen from (2.19) to be $\delta\nu(\mathbf{r})\sigma_0\mathbf{E}(\mathbf{r})$, thus (2.23) reduces to

$$dZ = -\int \sigma_0 \tilde{\mathbf{E}}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}) \delta\nu(\mathbf{r}) d\mathbf{r}. \tag{2.24}$$

But by definition,

$$dZ = \int \nabla Z(\mathbf{r}) \delta\nu(\mathbf{r}) d\mathbf{r}. \tag{2.25}$$

Then it is clear that the functional gradient of the impedance change is

$$\nabla Z(\mathbf{r}) = -\sigma_0 \tilde{\mathbf{E}}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}). \tag{2.26}$$

The gradient of the mean square error of the impedance change can be easily derived if the total electric fields $\mathbf{E}(\mathbf{r})$ and $\tilde{\mathbf{E}}(\mathbf{r})$ are known. The solution of the total fields can be best done by using the volume integral method because the Green's function appears directly in the formulation of the volume integral equation, and the solution of the hypothetical forward problem can be done in the same fashion as the solution of the normal forward problem.

Based on equation (2.26), a number of gradient based inversion methods can be used to restore the flaw conductivity distribution. These methods include the steepest descent algorithm, the conjugate gradient descent algorithm, and the Levenberg-Marquardt algorithm [27]. Various constraints can also be used to regularize the ill-conditioned inversion. However, the application of constraints cannot guarantee that the global minima can be found. Also, due to the long execution time needed to numerically solve the two forward problems in each iteration, this method is not very feasible for practical use.

CHAPTER 3. THE VOLUME INTEGRAL METHOD FOR EDDY CURRENT MODELING

Introduction

To develop signal and image processing techniques for two-dimensional eddy current inspection data, we must have a deep understanding of the underlying physical nature of eddy current measurements. Such an understanding not only gives insight on what techniques are appropriate for the eddy current problem, but also provides a basis for evaluating various processing techniques. The solution of the eddy current forward problem is also important for quantitative characterization of the shape and the size of the defect under investigation. The eddy current forward problem can be formally defined as the problem to predict the impedance change of a coil given the test geometry, the properties of the testpiece, the coil parameters, and the conductivity distribution in the flaw. Analytical solutions of the eddy current forward problem exist for simple test geometry and flaws with good symmetry properties. However, for flaws with irregular shapes in a complex geometry, an analytical solution usually is not available, so we must find a numerical solution. There have been several numerical models in the literature, e.g., the finite element method [7], the boundary element method [3, 4], and the volume integral method [2, 5, 6]. Among these models, the volume integral method has shown good potential due to its capability of modeling a three-dimensional flaw with arbitrary shape. It is also straightforward to apply the volume integral method if the Green's function of the given geometry is known.

In this chapter, we briefly review the volume integral method used in eddy current modeling. The volume integral method was originally used in geophysical induction studies and was introduced to the field of eddy current NDE by Dunbar [2]. The volume integral method transforms the governing diffusion equation of the eddy current problem to an integral equation, which is more suitable for numerical solution by using the Green's function for the given geometry. This volume integral equation can be discretized and transformed to a linear matrix equation which can be solved directly. In this chapter the volume integral method in half space geometry and thin plate geometry is discussed.

The Eddy Current Modeling Problem

Given a test geometry and the conductivity distribution in the flaw volume, the eddy current modeling problem is to find the impedance change of a coil driven by sinusoidal current and placed above the metal testpiece. We can derive the governing diffusion equation for this problem starting from the Maxwell equations

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{3.1}$$

$$\nabla \times \mathbf{H} = \mathbf{J}, \tag{3.2}$$

where $\mathbf{E}$ is the electric field intensity, $\mathbf{B}$ is the magnetic flux density, $\mathbf{H}$ is the magnetic field intensity, and $\mathbf{J}$ is the current density. In the frequency range of eddy current testing, the displacement current is very small, and thus can be neglected (quasistatic assumption). Taking the curl of (3.1) and substituting (3.2) into (3.1) by using the relation $\mathbf{B} = \mu \mathbf{H}$, we have the following equation

$$\nabla \times \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{J}}{\partial t}. \tag{3.3}$$

The current density in (3.3) contains two components

$$\mathbf{J} = \mathbf{J}^0 + \sigma \mathbf{E}, \tag{3.4}$$

where $\mathbf{J}^0$ is the current density due to the source coil, and the second term is the current induced by the total electric field, $\mathbf{E}$. Notice that in Cartesian coordinates

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}. \tag{3.5}$$

Since $\nabla \cdot \mathbf{E} = 0$ in metal (assume piece-wise continuous conductivity distribution in the flaw volume), we combine (3.5) and (3.3) with an $e^{i\omega t}$ time dependence to give the governing differential equation as

$$\nabla^2 \mathbf{E} - \gamma^2 \mathbf{E} = i\omega\mu \mathbf{J}^0, \tag{3.6}$$

where $\gamma^2 = i\omega\mu\sigma$. This is a diffusion equation, and the diffusive nature of eddy currents determines the resolution limit of eddy current testing. The source current density in this equation can be obtained from the incident field distribution. Analytical solutions exist for the incident field for an air-core coil above a half space or layered metal [28]. The fundamental problem of eddy current modeling is to solve equation (3.6) for the total field given the flaw conductivity distribution and the incident field.

Having solved for the total field, the impedance change of the coil next to a nonmagnetic metal can be obtained from the following reciprocity formula [1]

$$\Delta Z = -\frac{1}{I^2} \int_{V'} \delta\sigma(\mathbf{r}') \mathbf{E}^0(\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}') dv', \tag{3.7}$$

where I is the current in the coil, $\delta\sigma(\mathbf{r}')$ is the conductivity change, and V' is the flaw volume. This equation is particularly useful because it only requires the incident field and the total field to be evaluated within the flaw volume. It is not necessary to numerically compute the fields in the region of the coil.

The Volume Integral Method

In the following discussion of the volume integral method, we assume the geometry is either a half space of homogeneous metal or layered metal in which the flaw region is constrained within one layer. Consider a special solution of equation (3.6) when the current source is a point electric dipole placed in the layer containing the flaw

$$\nabla^2\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') - i\omega\mu\sigma_0\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = i\omega\mu\delta(\mathbf{r}-\mathbf{r}')\mathbf{I}, \tag{3.8}$$

where the solution of the above equation $\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')$ is the dyadic Green's function, $\delta(\mathbf{r}-\mathbf{r}')$ is the three-dimensional Dirac delta function, and $\mathbf{I}$ is the unit dyad. Given the Green's function for the test geometry, the solution of equation (3.6) can be expressed as a volume integral in the flaw region. To see this, rewrite equation (3.6) as

$$\nabla^2\mathbf{E} - i\omega\mu\sigma_0\mathbf{E} = i\omega\mu[\mathbf{J}^0 + (\sigma-\sigma_0)\mathbf{E}], \tag{3.9}$$

where $i\omega\mu\sigma_0\mathbf{E}$ has been subtracted from both sides of the equation. The term $(\sigma-\sigma_0)\mathbf{E} = \delta\sigma\mathbf{E}$ can be considered as an effective current dipole density at the source point due to the variation of the flaw conductivity from the host conductivity. We can combine equations (3.8) and (3.9), and express the total field as

$$\begin{aligned}\mathbf{E}(\mathbf{r}) &= \int \overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\cdot[\mathbf{J}^0(\mathbf{r}')+\delta\sigma(\mathbf{r}')\mathbf{E}(\mathbf{r}')]dv' \\ &= \mathbf{E}^0(\mathbf{r})+\int_{V'}\delta\sigma(\mathbf{r}')\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{E}(\mathbf{r}')dv',\end{aligned} \tag{3.10}$$

where $\mathbf{E}^0(\mathbf{r}) = \int \mathbf{G}(\mathbf{r},\mathbf{r}')\cdot\mathbf{J}^0(\mathbf{r}')dv'$. Given the Green's function and the incident field $\mathbf{E}^0$, we can solve equation (3.10) for the total field $\mathbf{E}$. This is usually done by discretizing equation (3.10). The flaw volume is divided into N smaller volume elements within which the electric field can be assumed constant. After discretization, equation (3.10) becomes

$$\mathbf{E}^0(\mathbf{r}) = \mathbf{E}(\mathbf{r}) + \sum_{i=1}^{N}\int_{V_i}\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\delta\sigma_i dv_i\cdot\mathbf{E}_i, \tag{3.11}$$

which can be further reduced to the following linear system

$$\mathbf{E}^0 = (\mathbf{I}-\mathbf{C})\mathbf{E}, \tag{3.12}$$

where $\mathbf{E}^0$ is incident field, $\mathbf{E}$ is the total field, $\mathbf{I}$ is the identity matrix and $\mathbf{C}$ is a matrix derived from the conductivity change and the integration of the Green's function in each element.

After equation (3.12) is solved and the total field is found, we can use a discretized version of equation (3.7) to compute the impedance change

$$\Delta Z = \sum_{i=1}^{N}\delta\sigma_i\mathbf{E}_i^0\cdot\mathbf{E}_i dv_i. \tag{3.13}$$

The Volume Integral Method for Half Space Geometry

To solve the volume integral equation (equation (3.11)), we must find the incident electric field and the Green's function for a given test geometry. The complexity of the

solutions for the incident field and the Green's function is dependent on the symmetry of the test geometry. One simple, but yet very useful geometry, is the half space geometry in which the flaw is located in a homogeneous metal half space. A thick metal plate can be modeled by the half space geometry if the thickness of the plate is much larger than the skin depth at the test frequency. The skin depth is defined as

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}}. \tag{3.14}$$

Incident field

To solve the eddy current forward problem using the volume integral method, we must find the incident field distribution for a given coil and test geometry. The incident field for a cylindrical air-core coil with rectangular cross-section placed above a homogeneous half space has been derived by Dodd and Deeds [28]. Their solution for the incident electric field is

$$E^0(r,z) = T\int_0^\infty \frac{I(\alpha)J_1(\alpha r)(e^{-\alpha l_1} - e^{-\alpha l_2})e^{-\alpha' z}}{\alpha(\alpha+\alpha')}d\alpha, \tag{3.15}$$

where

$$T = \frac{-j\omega\mu NI}{(l_2 - l_1)(r_2 - r_1)}, \tag{3.16}$$

$$I(\alpha) = \int_{r_1}^{r_2} rJ_1(\alpha r)dr, \tag{3.17}$$

and

$$\alpha' = \sqrt{\alpha^2 + j\omega\mu\sigma}. \tag{3.18}$$

In the above equations r_1 is the coil inner radius, r_2 is the coil outer radius, l_1 is the liftoff of the coil, l_2 is the distance between the metal surface and the coil upper surface, N is the number of turns of wire in the coil, I is the current in the coil, ω is the test frequency, and σ is the conductivity of the testpiece.

Green's function

To solve equation (3.11), we must also know the Green's function for the given test geometry. The Green's function for a homogeneous half space has been obtained analytically by Raiche and Coggon [29], and Beissner [30]. It includes two components, the primary or infinite medium component, and the secondary component which satisfies the boundary condition at the metal-air interface

$$G_{ij}(r,r') = G_{ij}^{P}(r,r') + G_{ij}^{S}(r,r'), \tag{3.19}$$

where the primary Green's tensor is given by

$$\begin{aligned} 4\pi\sigma G_{ij}^{P}(\mathbf{r},\mathbf{r}') &= \left(\frac{\partial^2}{\partial r_i \partial r_j} - \gamma^2\delta_{ij}\right)\frac{e^{-\gamma R}}{R} \\ &= \frac{e^{-\gamma R}}{R^3}\left[\left(3+3\gamma R+\gamma^2R^2\right)\frac{(r_i-r_i')(r_j-r_j')}{R^2} - \left(1+\gamma R+\gamma^2R^2\right)\delta_{ij}\right], \end{aligned} \tag{3.20}$$

where σ is the host conductivity, $\gamma = \sqrt{j\omega\mu\sigma}$, $i,j \in \{x,y,z\}$, δ_{ij} is the Kronecker delta function, and $R = \sqrt{(x-x')^2+(y-y')^2+(z-z')^2}$. From equation (3.20) it is clear that the primary Green's function is singular at $\mathbf{r} = \mathbf{r}'$.

The nine terms of the secondary Green's function G_{ij}^{S} are given by

$$4\pi\sigma G_{xx}^{S}=\left[\frac{2(x-x')^{2}}{r^{3}}-\frac{1}{r}\right]H_{1}-\frac{(x-x')^{2}}{r^{2}}H_{2}-\gamma^{2}H_{3}, \tag{3.21}$$

$$4\pi\sigma G_{xy}^{S}=\frac{(x-x')(y-y')}{r^{2}}\left[\frac{2H_{1}}{r}-H_{2}\right], \tag{3.22}$$

$$4\pi\sigma G_{xz}^{S}=\frac{-e^{-\gamma Q}}{Q^{5}}(x-x')(z+z')(3+3\gamma Q+\gamma^{2}Q^{2}), \tag{3.23}$$

$$G_{yx}^{S}=G_{xy}^{S}, \tag{3.24}$$

$$4\pi\sigma G_{yy}^{S}=\left[\frac{2(y-y')^{2}}{r^{3}}-\frac{1}{r}\right]H_{1}-\frac{(y-y')^{2}}{r^{2}}H_{2}-\gamma^{2}H_{3}, \tag{3.25}$$

$$4\pi\sigma G_{yz}^{S}=\frac{-e^{-\gamma Q}}{Q^{5}}(y-y')(z+z')(3+3\gamma Q+\gamma^{2}Q^{2}), \tag{3.26}$$

$$G_{zx}^{S}=-G_{xz}^{S}, \tag{3.27}$$

$$G_{zy}^{S}=-G_{yz}^{S}, \tag{3.28}$$

$$4\pi\sigma G_{zz}^{S}=\frac{-e^{-\gamma Q}}{Q^{3}}\left[\left(3+3\gamma Q+\gamma^{2}Q^{2}\right)\frac{(z+z')^{2}}{Q^{2}}-\left(1+\gamma Q+\gamma^{2}Q^{2}\right)\right], \tag{3.29}$$

where $r=\sqrt{(x-x')^{2}+(y-y')^{2}}$, $Q=\sqrt{(x-x')^{2}+(y-y')^{2}+(z+z')^{2}}$, and

$$H_{1}=\int_{0}^{\infty}\lambda\left(2-\frac{\lambda}{\nu}\right)e^{-\nu(z+z')}J_{1}(\lambda\gamma)d\lambda, \tag{3.30}$$

$$H_{2}=\int_{0}^{\infty}\lambda^{2}\left(2-\frac{\lambda}{\nu}\right)e^{-\nu(z+z')}J_{0}(\lambda\gamma)d\lambda, \tag{3.31}$$

$$H_{3}=\int_{0}^{\infty}\frac{\lambda}{\nu}\left(\frac{\nu-\lambda}{\nu+\lambda}\right)e^{-\nu(z+z')}J_{0}(\lambda\gamma)d\lambda, \tag{3.32}$$

are functions of γ and $(z+z')$, where $\nu = \sqrt{\lambda^2 + \gamma^2}$. The secondary Green's function is singular at $z+z'=0$.

The Volume Integral Method for Thin Plate Geometry

The half space geometry is appropriate for modeling surface breaking cracks or other near-surface flaws in a plate thick enough so that the second metal-air interface can be neglected, because in this case the diffusive fields cannot penetrate deep enough to the second surface. However, to model an OD crack on a steam generator tube, the half space geometry cannot be used. This is due to the test frequency usually being selected to make the skin depth comparable to the tube wall thickness so that the signal from an OD crack can be picked up by the coil. The direct modeling of the tube geometry is difficult because of lack of symmetry in the geometry. When the size of the coil is small compared to the tube, the tube internal surface near the coil is flat enough so that we can treat it as an infinite flat, thin plate. The thin plate geometry is also useful for many other applications of eddy current testing, e.g., the testing of hidden corrosion in lap-joints of aircraft.

Incident field

The incident field for a cylindrical air-core coil with rectangular cross-section placed above an infinite flat thin plate can also be derived from Dodd and Deeds' results in [28]. The expression for the incident electric field is

$$E^0(r,z)=T\int_{-\infty}^{\infty}\frac{I(\alpha)J_1(\alpha r)(e^{-\alpha l_1}-e^{-\alpha l_2})\left[(\alpha+\alpha')e^{-\alpha' z}+(\alpha'-\alpha)e^{-\alpha'(2H-z)}\right]}{\alpha\left[(\alpha+\alpha')^2-(\alpha'-\alpha)e^{-2\alpha' H}\right]}d\alpha, \qquad (3.33)$$

where T, $I(\alpha)$, α' are given in equations (3.16), (3.17) and (3.18), and H is the thickness of the plate.

Equation (3.33) reduces to equation (3.15) if we replace H with $+\infty$ in (3.33). Apparently the half space geometry is a special case of the thin plate geometry when the plate thickness is infinite.

Green's function

The Green's function for the thin plate case is more complicated than the Green's function for the half space geometry due to reflection terms between the two metal-air interfaces of the thin plate. Weaver [31] derived the Green's function for the geometry containing two homogeneous layers, and the thin plate geometry is a special case of the two-layer geometry. The Green's function for the thin plate geometry is based on the solution of two differential equations: the solution when the source is a vertical dipole and the solution when the source is a horizontal dipole. For the vertical dipole, the solution for the thin plate case can be derived by combining all the reflection terms of the Green's function for the half space geometry

$$G_{zz}^{th}=\sum_{n=-\infty}^{+\infty}\left[G_{zz}^{P}(x-x',\,y-y',\,z-z'+2nH)+G_{zz}^{S}(x-x',\,y-',\,z+z'+2nH)\right], \qquad (3.34)$$

where G_{zz}^{P} and G_{zz}^{S} are terms of the primary Green's function and the secondary Green's function for the half space geometry, respectively. Similarly,

$$G_{xz}^{th} = \sum_{n=-\infty}^{+\infty}\left[G_{xz}^{P}(x-x', y-y', z-z'+2nH) + G_{xz}^{S}(x-x', y-', z+z'+2nH)\right], \tag{3.35}$$

and

$$G_{yz}^{th} = \sum_{n=-\infty}^{+\infty}\left[G_{yz}^{P}(x-x', y-y', z-z'+2nH) + G_{yz}^{S}(x-x', y-', z+z'+2nH)\right]. \tag{3.36}$$

The solution for the horizontal dipole can be considered as the combination of two terms: the term for the Green's function in the half space geometry and an additional term

$$G_{ij}^{th} = G_{ij}^{0} + G_{ij}^{1}, \tag{3.37}$$

where $i \in \{x, y, z\}$, $j \in \{x, y\}$, G_{ij}^{0} is the term for the Green's function in the half space geometry, and G_{ij}^{1} is the additional term. The detail expressions for the additional terms can be found in Reference [31]. The computations of these additional terms require evaluations of integrations involving the zero order and the first order Bessel functions of the first kind.

CHAPTER 4. IMPLEMENTATION OF THE VOLUME INTEGRAL METHOD

Introduction

In the last chapter we discussed the volume integral method for the eddy current forward modeling problem. We also discussed analytical solutions of the incident fields and the Green's functions for half space geometry and thin plate geometry. The material discussed in the last chapter is the fundamentals for the implementation of the volume integral method. However, the content covered in the last chapter is not self-sufficient for an efficient and reliable implementation of the volume integral method. To actually implement the volume integral method, several problems must be carefully studied. One of the more important problems is how to accurately compute integrations involving Bessel functions. Another important problem is how to compute the volume integral for the Green's function in each volume element. In both cases, the integrations may not be stable due to the singular nature of the Bessel functions and the Green's function. In each step of the calculation we must find a computation method that is numerically tractable and efficient. In this chapter we discuss the problems associated with a practical implementation of the volume integral method. To start, we summarize the major steps of the volume integral method as follows:

1. Compute the incident field for a given test geometry and given coil parameters.
2. Compute the Green's function for the test geometry.

3. Calculate the system matrix by computing the volume integral of the Green's function in each volume element.
4. Invert the system matrix by LU decomposition.
5. Solve for the total field by back substitution.
6. Compute the impedance change by numerically integrating the impedance change volume integral.

Computation of the Incident Field

To calculate the incident field (equation (3.15) or equation (3.33)), we need to compute integrations with the first order Bessel function (equations (3.15), (3.17) and (3.33)). Direct numerical integration of these formulas are computationally intensive and intractable due to the singular nature of the Bessel function. Instead of direct integration, we can formulate the integrations in the form of first order Hankel (Fourier-Bessel) transforms and use an algorithm for fast Hankel transform to compute the integrations. The n th order Hankel transform is defined as

$$H_n(u) = \int_0^\infty rf(r)J_n(u)dr. \tag{4.1}$$

Comparing equations (3.15), (3.17) and (3.33) with equation (4.1), it is clear that these integrations are first order Hankel transforms.

There are many fast algorithms for computing Hankel transforms [32-36]. In our implementation we use an algorithm [35, 36] which does not require direct computation of the first order Bessel function of the first kind. This algorithm is based on the fact that the

Hankel transform can be considered as a series combination of the Chebyshev transform and the Fourier transform

$$H_n(u) = j^n F\{G_n\{f(r)\}\}, \tag{4.2}$$

where $F\{\ \}$ denotes the Fourier transform, and $G_n\{\ \}$ denotes the Chebyshev transform of order n, which is defined as [35]

$$G_n(y) = \int_{|y|}^{\infty} \frac{2T_n(y/r)f(r)dr}{\left(1-(y/r)^2\right)^{1/2}}, \tag{4.3}$$

where $T_n(x) = \cos(n\cos^{-1}(x))$ is a Cheybshev polynomial of the first kind, order n.

Based on equations (4.2) and (4.3), we can use the Simpson integration rule [37] to numerically compute the Chebyshev transform after using a change of variable $s^2 = r^2 - y^2$ in (4.3)

$$G_n(y) = 2\int_0^{\left(R^2-y^2\right)^{1/2}} f\left(\left(s^2+y^2\right)^{1/2}\right) T_n\left(y/\left(s^2+y^2\right)^{1/2}\right) ds. \tag{4.4}$$

After the numerical integration for the Chebyshev transform is finished, a Fast Fourier Transform (FFT) is then performed on the result of the numerical integration to obtain the result of the Hankel transform. This method is computationally efficient and stable. Only a moderate number of integration data points are necessary for an accurate result. One limit of the method is that the number of integration data points must be a power of 2 so that it can be processed by the FFT algorithm.

The incident field distribution for the half space geometry is computed using equations (3.15) - (3.18). The incident field distribution for the thin plate geometry is computed using equations (3.33) and (3.16) - (3.18). Notice that the incident field

expressions in equations (3.15) and (3.33) are two-dimensional functions of r and z. Therefore, a two-dimensional table is used to store the result of incident field computation. According to the modeling parameters, the incident field is computed for certain ranges of r and z. The range for z is usually decided by the vertical extent of the flaw region. The range for r is usually given by a constant multiple of the radius of the flaw region. At a certain depth z, two Hankel transforms must be computed to obtain the radial incident field distribution. In later modeling steps, the stored incident field distribution is read from the table and the vector electric field is transformed from a cylindrical coordinate system to an Euclidean coordinate system.

Computation of the Green's Function

Half space geometry

For half space geometry, the Green's function is comprised of two parts: the primary Green's function and the secondary Green's function. Computation of the primary Green's function for a given source point $\mathbf{r}'$ and a given field point $\mathbf{r}$ is straightforward by using equation (3.20). Computation of the secondary Green's function for a given source point $\mathbf{r}'$ and a given field point $\mathbf{r}$ is more complicated since it involves integrations of Bessel functions (equations (3.30) - (3.32)). As shown by Raiche and Coggon [29], the Hankel transforms of equations (3.30) - (3.32) can be evaluated analytically. Details of the analytical integrations can be found in Reference [29]. The results of the analytical integrations require evaluations of the zero order and the first order modified Bessel functions of the first kind I_0, I_1, and the second kind K_0, K_1.

The primary and secondary Green's function are computed using equations (3.20) - (3.32) and the analytical integration results in Reference [29]. The modified Bessel functions I_0, I_1, K_0, K_1 are approximated using power series expansions given in Reference [38]. Since the primary Green's function is a function of $x-x'$, $y-y'$, and $z-z'$, we compute it for various combinations of $x-x'$, $y-y'$, and $z-z'$, and store the results in a three-dimensional table. Due to the secondary Green's function being a function of $x-x'$, $y-y'$, and $z+z'$, we compute it for various combinations of $x-x'$, $y-y'$, and $z+z'$, and store the results in a three-dimensional table for later use.

Thin plate geometry

Computations for the Green's function in the thin plate geometry are more complicated. For a vertical dipole, we need to compute the summations in equations (3.34) - (3.36) based on our results for the Green's function in half space geometry. Apparently the infinite summations in equations (3.34) - (3.36) are not feasible. For any practical implementation, we can only use finite summations to approximate the infinite summations. The question is how to select a finite set of terms in the infinite summations to generate an accurate approximation, while minimizing the cost of computation. The answer to the question is based on the nature of eddy currents: the reflection terms in equations (3.34) - (3.36) decrease exponentially with n. The rate of decrease is defined by the skin depth at the test frequency. Therefore, by examining the values of the skin depth and the plate thickness we can find a number N, such that when $n > N$ or $n < -N$ the reflection terms are

practically zero. Only the reflection terms between $-N$ and N are used in the approximate finite summations.

For a horizontal dipole, the Green' function in the thin plate geometry is a combination of the Green's function in the half space geometry and an additional term (equation (3.37)). The additional term involves integrations with the zero order and the first order Bessel functions of the first kind. Following the approach used for the incident field computation, these integrations are also computed using the fast algorithm for zero order and first order Hankel transforms.

Computation of the System Matrix

Computation of the system matrix requires evaluation of integrations of the Green's function in rectangular volume elements. For both half space geometry and thin plate geometry, the Green's function can be separated into two parts: the primary Green's function which is a function of $x-x'$, $y-y'$, and $z-z'$, and the secondary Green's function which is a function of $x-x'$, $y-y'$, and $z+z'$. Due to the complexity of the secondary Green's function, its integration can only be performed numerically. The numerical integrations for the secondary Green's function are computed by using Gauss-Legendre quadrature [37]. The Legendre polynomial of order n is defined as

$$P_n(x)=\frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n. \tag{4.5}$$

The abscissas $x_1, x_2, \cdots, x_n$ in Gauss-Legendre quadrature are zeros of the Legendre polynomial

$$P_n(x_i) = 0, \quad i = 1, \cdots, n. \tag{4.6}$$

The corresponding weights $H_1, H_2, \cdots, H_n$ are given by

$$H_i = \frac{2(1-x_i)^2}{n^2 (P_{n-1}(x_i))^2}, \quad i = 1, \cdots, n. \tag{4.7}$$

To compute the abscissas and the weights, we need to find the zeros of the Legendre polynomial of order n. This is equivalent to solving a nonlinear equation which can usually be done by using the Newton's method. Details on numerical integrations with Gauss-Legendre quadrature can be found in Reference [37].

The volume integrals for the primary Green's function need more consideration because the primary Green's function is singular at $\mathbf{r} = \mathbf{r}'$. For the singular element (the volume element contains the point $\mathbf{r} = \mathbf{r}'$), we cannot use a numerical integration to compute the volume integral because it will lead to divergence. To make the integration for the singular element accurate, we must analytically integrate the singular part of the primary Green's function. Our approach is to use a Taylor series expansion of the primary Green's function to separate the singular part and the non-singular part. First we notice that

$$e^{-\gamma R} = 1 - \gamma R + \frac{\gamma^2 R^2}{2} - \frac{\gamma^3 R^3}{6} + O(R^4). \tag{4.8}$$

Substituting equation (4.8) into equation (3.20), we see that only the first and third terms in equation (4.8) result in singular terms. Therefore, we can separate the primary Green's function into two parts

$$4\pi\sigma G_{ij}^{P} = 4\pi\sigma G_{ij}^{P'} + 4\pi\sigma G_{ij}^{P''}, \tag{4.9}$$

where

$$4\pi\sigma G_{ij}^{P''} = \left(\frac{\partial^2}{\partial r_i \partial r_j} - \gamma^2 \delta_{ij}\right)\left(\frac{1}{R} + \frac{\gamma^2}{2} R\right) \tag{4.10}$$

contains singular terms in the primary Green's function, and $G_{ij}^{P'}$ contains non-singular terms in the primary Green's function. For the non-singular terms, we can numerically integrate them using Gauss-Lengendre quadrature. For the singular terms in equation (4.10), we must integrate them into a rectangular volume element analytically. The analytical integrations associated with the term $\left(\frac{\partial^2}{\partial r_i \partial r_j} - \gamma^2 \delta_{ij}\right)\frac{1}{R}$ have been derived in Reference [6] as

$$\iiint \frac{dxdydz}{\sqrt{x^2 + y^2 + z^2}} = f_1(x, y, z) + f_1(y, z, x) + f_1(z, x, y) + C, \tag{4.11}$$

where

$$f_1(x, y, z) = xy \log(\sqrt{x^2 + y^2 + z^2} + z) - \frac{x^2}{2} \arctan \frac{yz}{x\sqrt{x^2 + y^2 + z^2}}, \tag{4.12}$$

and

$$\iiint \frac{\partial}{\partial x} \frac{x}{\left(x^2 + y^2 + z^2\right)^{3/2}} dxdydz = \arctan \frac{yz}{x\sqrt{x^2 + y^2 + z^2}} + C, \tag{4.13}$$

$$\iiint \frac{\partial^2}{\partial x \partial y} \frac{1}{\sqrt{x^2+y^2+z^2}} dxdydz = \log(\sqrt{x^2+y^2+z^2}+z) + C. \tag{4.14}$$

We derived the analytical integrations associated with the term $\left(\frac{\partial^2}{\partial r_i \partial r_j} - \gamma^2 \delta_{ij}\right) R$ as

$$\begin{aligned} \iiint \frac{\partial^2}{\partial x^2} \sqrt{x^2+y^2+z^2}\, dxdydz &= xz \log(y+\sqrt{x^2+y^2+z^2}) \\ &+ xy \log(z+\sqrt{x^2+y^2+z^2}) - x^2 \arctan \frac{yz}{x\sqrt{x^2+y^2+z^2}} + C, \end{aligned} \tag{4.15}$$

$$\begin{aligned} \iiint \frac{\partial^2}{\partial x \partial y} \sqrt{x^2+y^2+z^2}\, dxdydz &= \frac{z\sqrt{x^2+y^2+z^2}}{2} \\ &+ \frac{x^2+y^2}{2} \log(z+\sqrt{x^2+y^2+z^2}) + C. \end{aligned} \tag{4.16}$$

The analytical result for the integration $\iiint \sqrt{x^2+y^2+z^2}\, dxdydz$ is not shown here because it is not singular and can be integrated numerically.

Solution of the Linear System

After the incident field is computed and the system matrix is obtained, we can form a system of linear equations as shown in equation (3.12). The vector containing the incident field is obtained by numerically integrating the incident field within each rectangular volume element. The solution of equation (3.12) is usually done by using LU decomposition with Gauss-Jordon elimination [37]. A square matrix $\mathbf{A}$ can be decomposed into a lower-triangular matrix $\mathbf{L}$ and an upper-triangular matrix $\mathbf{U}$ if it is not singular

$$\mathbf{A} = \mathbf{LU}. \tag{4.17}$$

The Gauss-Jordon elimination algorithm for LU decomposition iteratively performs row operations to transform the original matrix into an upper-triangular matrix, while the lower-triangular matrix is defined by the coefficients of the row operations.

To improve numerical stability, an operation called pivoting is usually used in the Gauss-Jordon elimination. The pivoting operation finds the number with the largest absolute value in the current column of elimination and exchanges the current row with the row containing the largest pivot. By performing pivoting, the numbers used in the elimination are divided by the largest number available. This avoids divisions of small numbers which may lead to erratic results due to the limited accuracy of any practical computing device.

Due to all quantities in equation (3.12) being complex numbers, a complex version of LU decomposition must be used. The complex version of the LU decomposition algorithm is basically the same as the real value LU decomposition algorithm presented in Reference [37] except all real operations are replaced with complex operations.

Since many elements in the system matrix are very small, the system matrix can be considered a sparse matrix. Based on the sparse nature of the system matrix, we can use a sparse matrix implementation of the LU decomposition algorithm to reduce the computational complexity of inverting the matrix. The sparse matrix LU algorithm is based on a single link list data structure. Each row of the sparse matrix is stored in one single link list. The sparse matrix LU algorithm not only reduces the computations involved in matrix inversion, but also greatly reduces the memory requirement.

Computation of the Impedance Change

After the system matrix is inverted by the LU decomposition, we can compute the total field by using back substitution. Details on the back substitution algorithm can be found in Reference [37]. If pivoting is used in the Gauss-Jordon elimination algorithm, components in the incident field vector must be reordered according to the pivoting sequence to ensure a correct result. Corresponding to the sparse matrix LU decomposition algorithm, we have implemented a sparse matrix back substitution algorithm which is also based on the single link list data structure. After the total field is solved for every volume element, the impedance change of the coil can be obtained using equation (3.13). This process is usually repeated for different offsets between the coil center and the flaw center. The number of offsets and the stepsizes are defined by the scan plan used in the modeling program.

CHAPTER 5. WAVELET EXPANSION IN THE VOLUME INTEGRAL METHOD

Introduction

In the last two chapters we discussed the theory and the implementation of the volume integral method in half space geometry and thin plate geometry. Although the volume integral method is capable of modeling three-dimensional flaws, the computation involved in modeling large three-dimensional flaws is usually far beyond what can be achieved by conventional computing sources because a large number of volume elements must be used. If the number of volume elements is N, the memory requirement for storing the system matrix is proportional to N^2, and the number of operations required to invert the system matrix is on the order of N^3. Also, the accuracy of the method is related to the size of the volume elements. In metal the electric field change is closely related to the skin depth; thus it is necessary for the dimension of each volume element to be much smaller than the skin depth to accurately model the field distribution. For a large three-dimensional flaw (in units of skin depth), the number of elements required will be on the order of thousands. Such a large number of elements causes a very high, sometimes almost unattainable computational requirement for the computing resource. Therefore, to apply the volume integral method to large flaws, we must find a way to reduce the computational complexity.

In this chapter we apply a wavelet expansion to the volume integral method to reduce the computational complexity of the method. Part of the content in this chapter has been

presented in Reference [39] and [40]. Results from the wavelet method have been compared with results from a layer approximation and experimental results [12]. In this chapter we give a comprehensive discussion of the wavelet expansion method, with both theoretical derivations and implementation details. Utilizing this method, a new linear system is obtained by using wavelet expansion in a Galerkin method to solve the governing volume integral equation. By using a proper threshold, a sparse system matrix can be generated which results in a large savings in execution time and memory requirements.

To evaluate the performance of the wavelet method, we present several numerical examples with the Haar wavelet [41] and the Daubechies' compactly supported wavelet [42] with a periodic extension. The numerical examples show how the wavelets change the sparsity of the system matrix. We also observed an interesting relationship between the physical nature of the eddy current problem and the structure of the system matrix. To evaluate the accuracy lost in the process of the wavelet transform and threshold, we conducted an experiment on a simulated corrosion pit underneath a thin aluminum plate. The experimental results are compared with results given by the wavelet method. We observed that a very high compression rate can be achieved with only a small loss in accuracy.

Introduction to Wavelet Theory

Based on a rigorous mathematical foundation, wavelets and multiresolution analysis have been widely applied in many engineering disciplines such as signal processing, data compression and pattern recognition. The unique time-frequency (or spatial-wave number) localization property of wavelets provides a new tool with which we can form a

multiresolution view of the signal under investigation. We give a brief introduction to wavelet theory and multiresolution analysis, with some emphasis on the orthonormal wavelet basis and periodic wavelets. The discussion is limited to the fundamental theory and the aspects of wavelets that are essential for the completeness of the discussion in this chapter. For detailed discussion on wavelet theory, one may refer to the books of Chan [41], Daubechies [42], and Chui [43], and the papers by Vetterli [44], Heil and Walnut [45], and Mallat [46].

Wavelets and multiresolution analysis

Wavelets are the families of basis functions which are derived from a common function - the mother wavelet that is represented by

$$\psi^{a,b}(x) = |a|^{-1/2}\,\psi\left(\frac{x-b}{a}\right), \tag{5.1}$$

where a and b are the scaling and translation indices, respectively. The above expression of a wavelet basis suggests a multigrid representation. Unlike other basis functions, such as the Fourier basis, the wavelet basis has two indices in both time and frequency domains. It is well known that, because Fourier basis functions are periodic functions, they are localized only in the frequency domain, but not in the time domain. The uncertainty principle imposes limits on the simultaneous localization in both time and frequency domains. Therefore, for any single basis function, we cannot localize it in both time and frequency domains. Although we still have to follow the uncertainty principle for a wavelet basis, we are able to localize some basis functions in the time domain and others in the frequency domain by selecting proper values of the scaling index and translation index for each basis function.

Therefore, we are able to form a multiresolution representation of the linear space under investigation by using a wavelet basis.

The mathematical foundation of the construction of multiresolution analysis starts from the definition of the scaling function, which is also called the father wavelet. A function $\phi(x)$ in $L^2(R)$ (finite energy function) is called a scaling function if it generates a nested sequence of subspaces

$$\{0\} \leftarrow \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \rightarrow L^2(R) \tag{5.2}$$

and satisfies the dilation equation

$$\phi(x) = \sum_k c_k \phi(ax - k), \tag{5.3}$$

with $\{c_k\}$ being a finite energy sequence and a being any rational number. A common selection is $a = 2$, which is the case for octave scales, and the subspaces V_j are generated by $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$. Since, for any scale j, we have $V_j \subset V_{j+1}$, there exists a unique orthogonal complementary subspace W_j of V_j in V_{j+1}. This subspace W_j is called "wavelet subspace" and is generated by $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, where $\psi(x)$ is the mother wavelet. Based on this relationship, the mother wavelet can be derived from the corresponding scaling function using

$$\psi(x) = \sum_k (-1)^k c_{1-k} \phi(2x - k). \tag{5.4}$$

The scaling function exhibits low-pass filter characteristics in the sense that $\hat{\phi}(0) = 1$, where $\hat{\phi}(\xi)$ denotes the Fourier transform of function $\phi(x)$. On the other hand, the wavelet function exhibits band-pass filter characteristics in the sense that $\hat{\psi}(0) = 0$. Because any

subspace V_{j+1} is the direct sum of the subspaces V_j and W_j , it follows that V_j contains the low resolution components in V_{j+1}, and W_j contains the high resolution components in V_{j+1}. Since $\{V_j\}$ is a nested sequence ranging from the subspace with the lowest resolution to the subspace with the finest detail, we therefore have a multiresolution decomposition of $L^2(R)$ in which V_j contains the smoothed part at resolution level j, and W_j contains the detailed part at resolution level j. A multiresolution decomposition of $L^2(R)$ also possesses the property that any function in $L^2(R)$ can be approximated as closely as possible by a function that belongs to at least one of the subspaces V_j.

Orthonormal wavelet basis

The wavelets $\{\psi_{j,k}\}$ form an orthonormal basis if

$$\langle \psi_{j,k}, \psi_{l,m} \rangle = \delta_{j,l}\delta_{k,m}; \text{ for all } j,k,l,m \in Z \tag{5.5}$$

where $\langle f, g \rangle$ is the inner product operation of functions $f(x)$ and $g(x)$ in $L^2(R)$, and is defined by

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(x)g(x)dx, \tag{5.6}$$

Z is the set of integers and $\delta_{p,q}$ is the Kronecker delta function.

The significance of an orthonormal wavelet basis lies in the fact that it is a non-redundant representation of the original signal. The construction of a multiresolution

decomposition of $L^2(R)$ requires the subspaces $\{W_j\}$ be orthogonal to each other. Given any function $f(x)$ in $L^2(R)$, we can approximate it by using the orthonormal wavelet basis

$$f(x) = c + \sum_j \sum_k a_{j,k} \psi_{j,k}(x), \tag{5.7}$$

where c is a constant, and the expansion coefficients $a_{j,k}$ can be obtained by projecting $f(x)$ onto the subspace expanded by $\psi_{j,k}(x)$ and

$$a_{j,k} = \langle f(x), \psi_{j,k}(x) \rangle. \tag{5.8}$$

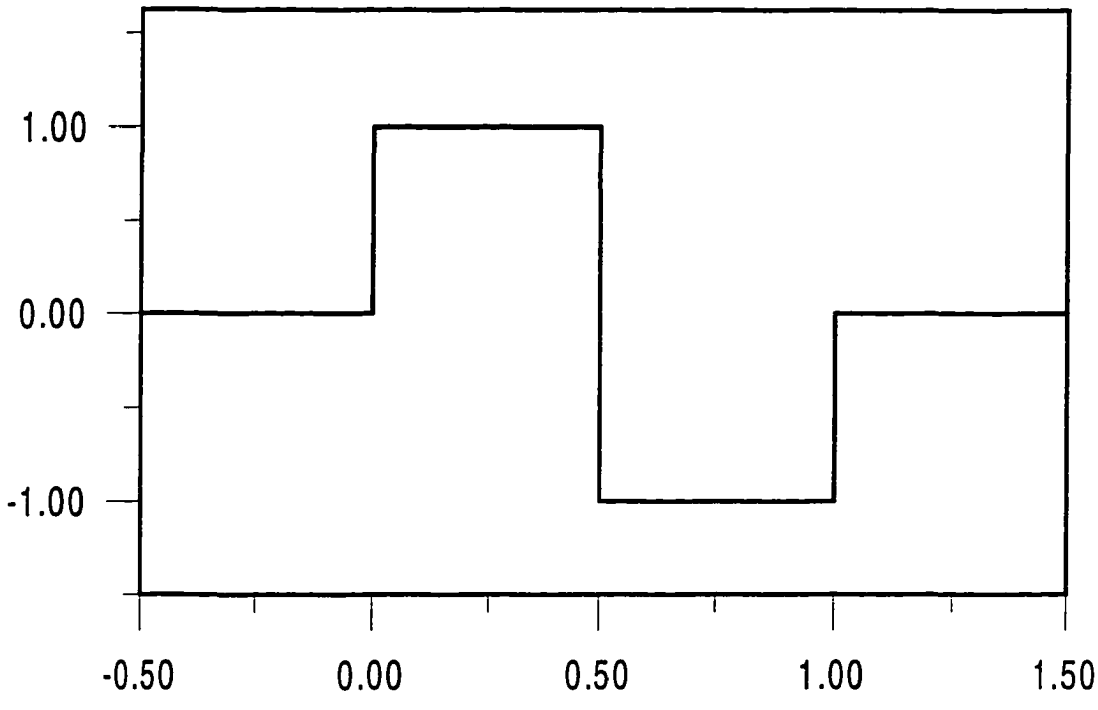
It is obvious that not every function that satisfies equation (5.3) can be used as the scaling function to construct an orthonormal wavelet basis. To construct an orthonormal wavelet basis, some additional conditions must be satisfied so that the form of scaling function is constrained. There are two fundamental conditions that define an orthonormal wavelet basis: the approximation condition and the orthogonality condition. The approximation condition comes from the question: To what degree $p-1$ can the polynomials $1, x, x^2, \cdots, x^{p-1}$ be reproduced exactly by the approximating functions, which, in our case, are the scaled and translated versions of the solution of equation (5.3)? If a polynomial of order $p-1$ can be approximated exactly, then the approximation error for an arbitrary function is of the order h^p, where h is the grid separation. From approximation theory, we know that if ϕ has an approximation error of order h^p, its Fourier transform $\hat{\phi}$ must have zeros of order p at all points $\xi = 2\pi n, n \in Z$ (except at $\xi = 0$) [42]. This gives us the following constraint on the recursion coefficients in equation (5.3),

$$\sum_k (-1)^k k^m c_k = 0, \quad m = 0, 1, \cdots, p-1. \tag{5.9}$$

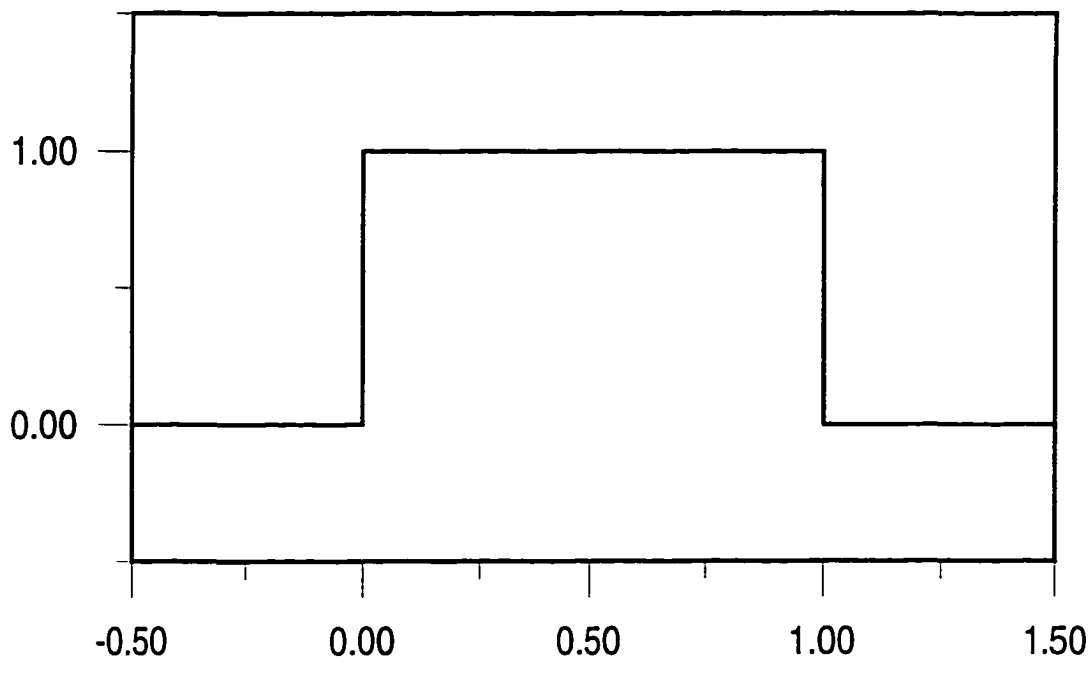
The orthogonality condition comes from the requirement that the wavelets defined in equation (5.4) must be orthogonal to each other. This gives another constraint on the recursion coefficients [42] of

$$\sum_k c_k c_{k-2m} = 2\delta_{0m}. \tag{5.10}$$

The objective in constructing orthonormal wavelets is to satisfy conditions of equations (5.9) and (5.10). The simplest orthonormal wavelet is the Haar wavelet with $p = 1$. The corresponding scaling function is the box function. Fig. 5.1 illustrates the Haar wavelet and the corresponding scaling function. The Haar wavelet has a very compact support - as small as the grid separation. This makes it the simplest wavelet that can be applied to function approximation and expansion. However, since it has $p = 1$, it is rather inefficient in approximating smooth functions. Another group of compactly supported orthonormal wavelets has been developed by Daubechies, which has an accuracy of $p = 1, 2, 3, \cdots$ and there are $2, 4, 6, \cdots$ nonzero coefficients. Fig. 5.2 illustrates the Daubechies wavelet with $p = 2$ and the corresponding scaling function. The selection of p is a tradeoff resulting from the uncertainty principle: a large p gives better approximation of smooth functions as well as a larger support for the corresponding wavelet basis. In real applications a compromise must be made on p according to the nature of the problem under investigation.

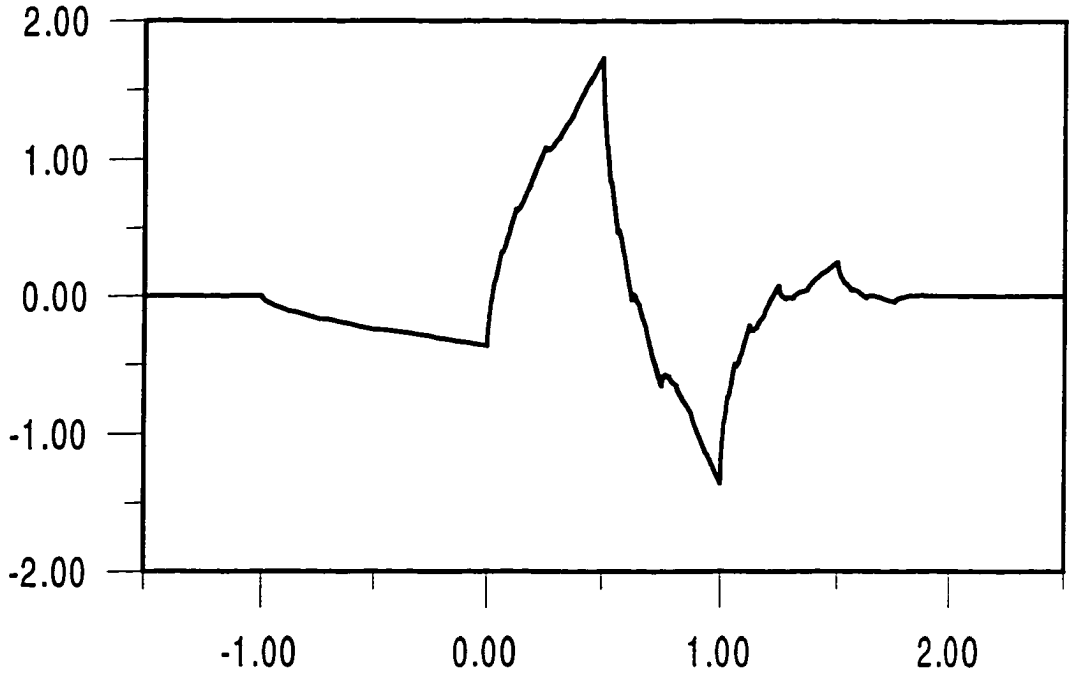


(a) Haar wavelet

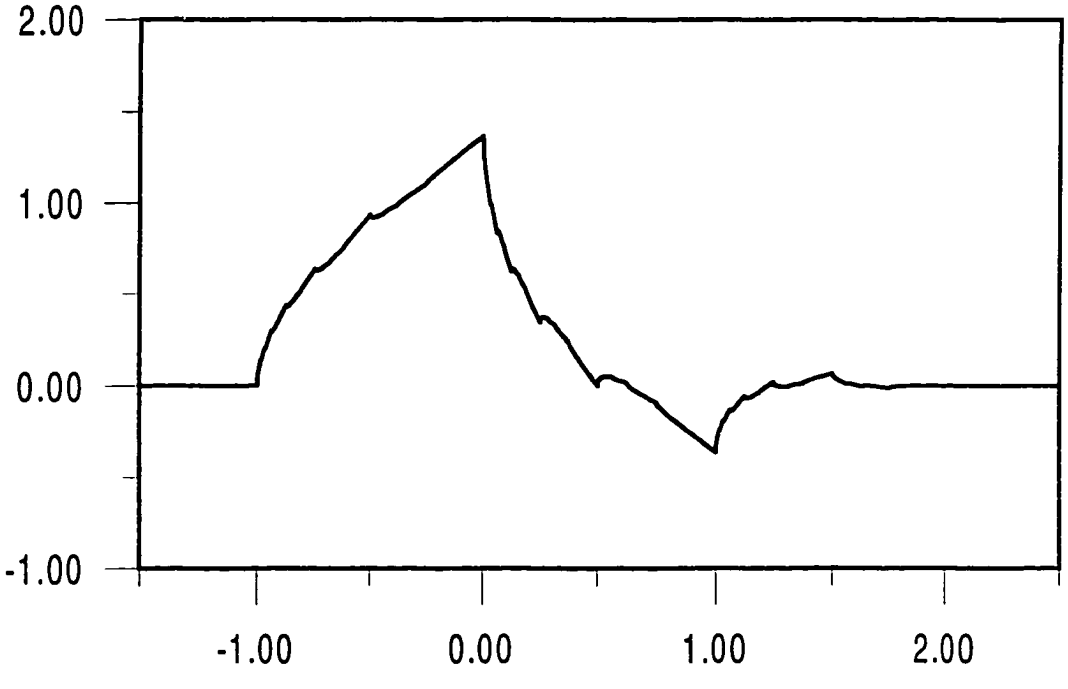


(b) The box function

Fig. 5.1. Haar wavelet and its scaling function - the box function.



(a) The Daubechies compactly supported wavelet with $p = 2$



(b) The corresponding scaling function

Fig. 5.2. The Daubechies compactly supported wavelet with $p = 2$ and its scaling function.

Periodic wavelets

In the section above we discussed orthonormal wavelets on the real line. For the purpose of solving an integral equation in a bounded region, it is more convenient if we use wavelets that are periodic on the real line, as we will explain later. Given an orthonormal multiresolution analysis with scaling function $\phi(x)$ and wavelet $\psi(x)$ on the real line, the periodic, orthonormal wavelets in $[0,1]$ can be defined as [42]

$$\phi_{j,k}^{per}(x)=\sum_{n\in Z}\phi_{j,k}(x+n), \tag{5.11}$$

$$\psi_{j,k}^{per}(x)=\sum_{n\in Z}\psi_{j,k}(x+n) \tag{5.12}$$

with

$$\phi_{j,k}(x)=2^{j/2}\phi(2^j x-k), \tag{5.13}$$

$$\psi_{j,k}(x)=2^{j/2}\psi(2^j x-k) \tag{5.14}$$

where $j,k,n\in Z$. It can be shown [42] that $\phi_{0,0}^{per}=1$ and $\{\psi_{j,k}^{per}\,|\,j=0,1,2,\cdots,k=1,\cdots,2^j-1\}$ constitute a periodic, orthonormal basis in $[0,1]$.

For any periodic function in $L^2(R)$ $[0,1]$, an approximation of this function can be defined as the projection at the resolution level j

$$f(x)\approx P_j f(x)=\sum_{k=0}^{2^j-1} a_{j,k}\psi_{j,k}^{per}(x), \tag{5.15}$$

where $a_{j,k}$ is the inner product of $f(x)$ and $\psi_{j,k}^{per}(x)$ in $[0,1]$. P_j is called the projection operator onto the subspace V_j^{per} - the subspace formed by the linear combination of basis functions $\{\phi_{j,k}^{per}(x)|k\in Z\}$. Since $\phi_{j,k}^{per}(x)$ has a typical length scale of 2^{-j}, V_j^{per} can be

referred to as the subspace of all the functions that have no detail with length scales smaller than 2^{-j}. Therefore, the subspace V_{j-1}^{per} is a subset of the subspace V_j^{per}. According to wavelet theory, every function in $L^2(R)[0,1]$ can be approximated to arbitrary resolution by its projection in V_j^{per} and the projection will converge to the original function as j approaches ∞.

If the orthonormal wavelet and scaling function in equations (5.13) and (5.14) are compactly supported, the infinite summations in equations (5.11) and (5.12) reduce to finite summations. The periodic, orthogonal wavelet $\psi_{0,0}^{per}(x)$ constructed from the Daubechies compactly supported wavelet with $p=2$ on the real line is shown in Fig. 5.3.

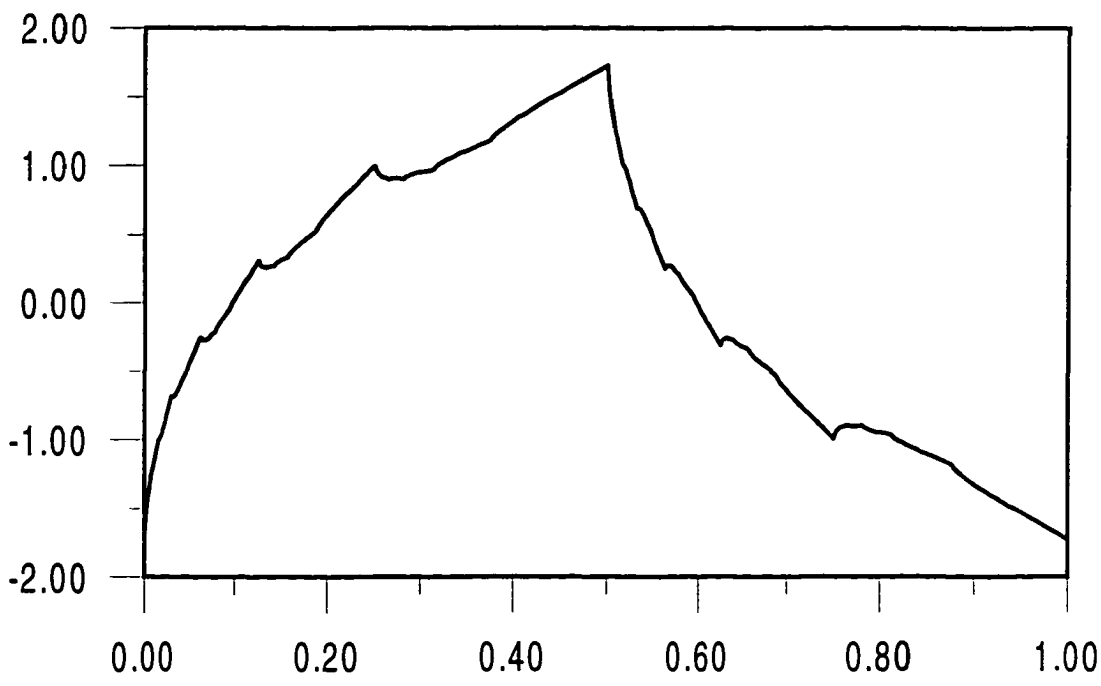


Fig. 5.3. The periodic wavelet basis $\psi_{0,0}^{per}(x)$ constructed from the Daubechies compactly supported wavelet with $p=2$.

Wavelet Expansion in the Solution of the Volume Integral Equation

In this section we apply the wavelet basis discussed in the last section to the solution of the three-dimensional volume integral equation introduced in Chapter 3. Wavelets have been applied to the solutions of integral equations [47-49]. To simplify the mathematical derivation, we consider the one-dimensional case first and derive the new linear system equations by using the Galerkin method. Then we extend the result to the three-dimensional case. The discussion will be based on using the wavelet basis on the real line. The modification of the method when using periodic wavelets will also be considered.

One-dimensional case

Consider the following one-dimensional simplification of the problem defined in equation (3.10)

$$E^0(x) = E(x) - \int_\Omega G(x,x')\delta\sigma(x')E(x')dx', \tag{5.16a}$$

$$E(x) = 0, \qquad x \notin \Omega, \tag{5.16b}$$

where $\Omega = [0, L]$ is the solution region which contains the flaw. The explicit boundary condition of equation (5.16b) is the result of using the wavelets on the real line, since some of the wavelet basis functions used for the expansion of the unknown total field $E(x)$ in the solution region Ω do not vanish outside Ω. Then we use the Galerkin method to approximate $E(x)$ in Ω with an orthonormal wavelet basis on the real line giving

$$E(x) = \frac{a_0}{\sqrt{L}} + \sum_{j=j_1}^{j_2} \sum_{k=k_1(j)}^{k_2(j)} a_{jk}\psi_{j,k}(x), \qquad x \in \Omega, \tag{5.17}$$

where a_0 and a_{jk} are the expansion coefficients and $\sqrt{L}$ is a normalization factor. The expansion of $E(x)$ uses wavelets from resolution level j_1, which corresponds to the lowest resolution component of $E(x)$, to the resolution level j_2, which corresponds to the desired resolution for the solution. Notice that equation (5.17) is only satisfied when $x \in \Omega$. Letting

$$p(x) = \begin{cases} 1 & x \in \Omega \\ 0 & x \notin \Omega, \end{cases} \tag{5.18}$$

we can express the total field on the real line as

$$E(x) = \frac{a_0 p(x)}{\sqrt{L}} + \sum_{j=j_1}^{j_2} \sum_{k=k_1(j)}^{k_2(j)} a_{jk} \psi_{j,k}(x) p(x). \tag{5.19}$$

To simplify the expression, we can combine the scaling index and translation index of the basis functions into one index. Let

$$\psi_0(x) = \frac{p(x)}{\sqrt{L}} \tag{5.20}$$

and suppose the number of basis functions used in the summation of equation (5.19) is $N-1$, the total field can be expressed as

$$E(x) = \sum_{i=0}^{N-1} a_i \psi_i(x) p(x), \tag{5.21}$$

where $\{a_i\}$, $\{\psi_i(x)\}$ are derived from $\{a_{jk}\}$, $\{\psi_{j,k}(x)\}$ by combining the indices. Substituting equation (5.21) into equation (5.16a) gives

$$E^0(x) = \sum_{i=0}^{N-1} a_i \psi_i(x) p(x) + \int_{\Omega} G(x,x') \delta\sigma(x') \sum_{i=0}^{N-1} a_i \psi_i(x') p(x') dx'. \tag{5.22}$$

Since $\delta\sigma(x') = 0$ if $x \notin \Omega$, we can rewrite the right hand side of equation (5.22) as

$$E^0(x) = \sum_{i=0}^{N-1} a_i \psi_i(x) p(x) + \sum_{i=0}^{N-1} a_i \langle G(x,x')\delta\sigma(x'), \psi_i(x') \rangle. \tag{5.23}$$

Taking the inner product of both sides of equation (5.23) with $\psi_j(x)$, $j = 0, 1, \cdots, N-1$ gives

$$\langle E^0(x), \psi_j(x) \rangle = \sum_{i=0}^{N-1} a_i \langle \psi_i(x) p(x), \psi_j(x) \rangle + \sum_{i=0}^{N-1} a_i \langle \langle G(x,x')\delta\sigma(x'), \psi_i(x') \rangle, \psi_j(x) \rangle, \tag{5.24}$$

$$j = 0, 1, \cdots, N-1.$$

These linear equations can be written in more compact form as

$$\sum_{i=0}^{N-1} a_i (D_{ij} + C_{ij}) = b_j, \quad j = 0, 1, \cdots, N-1 \tag{5.25}$$

where $b_j = \langle E^0(x), \psi_j(x) \rangle$ is the wavelet expansion coefficient of the incident field, $C_{ij} = \langle \langle G(x,x')\delta\sigma(x'), \psi_i(x') \rangle, \psi_j(x) \rangle$ is the two-dimensional expansion coefficient of the function $G(x,x')\delta\sigma(x')$, and $D_{ij} = \int_{\Omega} \psi_i(x)\psi_j(x)dx$. Given the Green's function for the test geometry, the incident field distribution and the conductivity distribution in the flaw region, b_j and C_{ij} can be easily obtained by using the fast wavelet transform algorithm, and D_{ij} is derived directly from the wavelet basis selected.

Equation (5.25) is the general form of the linear system equations for the solution of equation (5.16a) after using an orthonormal wavelet basis on the real line. The reason why D_{ij} does not reduce to the Kronecker delta function is that the basis functions $\{\psi_i(x)\}$ used to approximate the total field are not always orthogonal to each other in Ω. Suppose the support of $\psi_i(x)$ is S_i, and the support of $\psi_j(x)$ is S_j, then we have

$$D_{ij} = \delta_{ij} \qquad \text{if} \quad S_i \cap S_j \subseteq \Omega. \tag{5.26}$$

Therefore, if any part of the overlapped region of S_i and S_j are outside of Ω, the value of D_{ij} must be calculated explicitly. This is another disadvantage of using wavelets on the real line in the solution of equation (5.16a) in a bounded region. A special case of equation (5.25) exists when we use the Haar wavelet to approximate the unknown total field. In this case, we are able to express the total field in Ω by using basis functions all lying in Ω because the Haar wavelet has a support as small as the grid separation, and $D_{ij} = \delta_{ij}$ because $S_i \subseteq \Omega., i = 0, 1, \cdots, N-1$. Therefore, in the case of the Haar basis, there is no problem at the boundary of the solution region and the boundary condition of equation (5.16b) is not necessary. However, since the Haar basis is not efficient in approximating smooth functions ($p = 1$), the resultant system matrix may not be as sparse as that from a smoother wavelet.

Extension to the three-dimensional case

The above result can be extended to the three-dimensional case by a similar, but somewhat lengthy derivation. The three-dimensional volume integral problem is given by

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) + \int_{V'} \delta\sigma(\mathbf{r}')\bar{\mathbf{G}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}')dv', \tag{5.27a}$$

$$\mathbf{E}(\mathbf{r}) = \mathbf{0}, \qquad \mathbf{r} \notin V'. \tag{5.27b}$$

We can construct an orthonormal basis for $L^2(R^3)$ by using the tensor product functions generated by three one-dimensional orthonormal wavelet bases

$$\psi_{j_x,k_x;j_y,k_y;j_z,k_z}(x,y,z) = \psi_{j_x,k_x}(x)\psi_{j_y,k_y}(y)\psi_{j_z,k_z}(z), \tag{5.28}$$

where the wavelets in x, y and z directions are scaled independently.

By a similar formulation and combination of indices as we used in the section above, we can express the total vector field as

$$\mathbf{E}(\mathbf{r}) = \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}\sum_{k=0}^{N_z-1}\mathbf{a}_{ijk}\psi_{i;j;k}(x,y,z)p_x(x)p_y(y)p_z(z), \tag{5.29}$$

where $\mathbf{a}_{ijk}$ is the vector expansion coefficient for the three-dimensional total vector field $\mathbf{E}(\mathbf{r})$.

Substituting equation (5.29) into equation (5.27a) and using the same procedure as in the section above, we can obtain the linear system equations for the three-dimensional case as

$$\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}\sum_{k=0}^{N_z-1}\mathbf{a}_{ijk}\cdot(\overline{\mathbf{D}}_{ijklmn}+\overline{\mathbf{C}}_{ijklmn}) = \mathbf{b}_{lmn} \tag{5.30}$$

$$l = 0, 1, \cdots, N_x - 1, \quad m = 0, 1, \cdots, N_y - 1, \quad n = 0, 1, \cdots, N_z - 1,$$

where $\mathbf{b}_{lmn}$ is the vector expansion coefficient for the incident vector field, $\overline{\mathbf{C}}_{ijklmn}$ is the expansion coefficient of the six-dimensional function $\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\delta\sigma(\mathbf{r}')$ and $\overline{\mathbf{D}}_{ijklmn}$ is a 3 by 3 matrix with 9 identical elements, and each element can be expressed as

$$d_{ijklmn} = \int_{V'}\psi_{i;j;k}(x,y,z)\psi_{l;m;n}(x,y,z)dv' \tag{5.31}$$

Extension to the periodic wavelet case

As seen before, using wavelets on the real line in the solution of integral equations like equation (5.16a) has two major disadvantages. First, the boundary condition of equation (5.16b) must be explicitly enforced. This is a very undesirable requirement in applying this

method since the total field generally does not vanish outside the flaw region. To satisfy (5.16b), we must enlarge the solution region Ω so that outside Ω the total field is very small and can be neglected. This may result in a solution region which is much larger than the flaw region, and thus reduce the usefulness of this method. Another major disadvantage is that because some of the wavelet basis functions used to approximate the total field are not orthogonal in the solution region, the term D_{ij} does not reduce to delta functions and must be calculated before the solution of equation (5.25). These disadvantages arise from the fact that wavelet basis functions on the real line used to approximate the total field do not form an orthogonal set in the solution region Ω. The periodic wavelet introduced in the last section, on the other hand, can be used to avoid these difficulties because it can be made orthonormal in a bounded region Ω. Using an orthonormal, periodic wavelet basis $\left\{\psi_{j,k}^{per}(x)\right\}$ defined in $\Omega = [0, L]$, we can expand the total field in Ω as

$$E(x) = \sum_{i=0}^{N-1} a_i \psi_i^{per}(x), \qquad x \in \Omega \tag{5.32}$$

by combining indices. Notice that the right hand side of equation (5.32) denotes a periodic function with a period equal to L and whose value is equal to the total field in region Ω. Substituting equation (5.32) into equation (5.16a), we have

$$E^0(x) = \sum_{i=0}^{N-1} a_i \psi_i^{per}(x) + \int_{\Omega} G(x,x')\delta\sigma(x') \sum_{i=0}^{N-1} a_i \psi_i^{per}(x')dx', \qquad x \in \Omega \tag{5.33}$$

Projecting both sides of equation (5.33) into the subspaces expanded by $\psi_j^{per}(x)$, $j = 0, 1, \cdots, N-1$, we have the following equations

$$\int_{\Omega} E^0(x)\psi_j^{per}(x)dx = \sum_{i=0}^{N-1} a_i\delta_{ij} + \sum_{i=0}^{N-1} a_i \iint_{\Omega\Omega} G(x,x')\delta\sigma(x')\psi_i^{per}(x')\psi_j^{per}(x)dx'dx, \quad (5.34)$$

$$j = 0, 1, \cdots, N-1,$$

which can be rewritten in the form of a matrix equation

$$\sum_{i=0}^{N-1} a_i(\delta_{ij} + C_{ij}) = b_j, \quad j = 0, 1, \cdots, N-1 \quad (5.35)$$

where b_j are the periodic wavelet expansion coefficients of the incident field and C_{ij} are the periodic wavelet expansion coefficients of $G(x,x')\delta\sigma(x')$. The use of periodic wavelets in the solution of equation (5.16a) gives a group of well formed linear equations: all coefficients in equation (5.35) can be obtained by using the fast wavelet transform algorithm of Mallat [46]. Also, since the boundary condition equation (5.16b) is not necessary, we can use a solution region which has the same size as the flaw region.

Despite its advantage over the wavelet on the real line, a potential disadvantage of using periodic wavelets exists due to the fact that the periodization of the total field may introduce an edge effect. Since the values of the total field at the two ends of the solution region generally are not equal, there are discontinuous points at the boundaries of each period. This discontinuity introduces some large expansion coefficients for the total field and makes the system matrix less sparse. However, experience has shown that this edge effect generally does not have a large influence on the overall performance of the wavelet method.

Numerical Results and Comparison with Experimental Results

To evaluate the validity and performance of the wavelet method, we implemented the method and used the problem of hidden corrosion in an aluminum plate as a test case. In this section, we first describe the experimental setup for the model problem used as the test case for the wavelet method. Then we give some implementation considerations of the wavelet method. We also give the numerical results obtained by using the Haar wavelet and the periodic wavelet constructed from the Daubechies compactly supported wavelets with $p = 2$, and compare the results from the wavelet method with the experimental result. The experimental results presented in this chapter were obtained by John Moulder of the center for NDE at Iowa State University.

The experiment

The test case is to determine the change in the impedance of a right-cylindrical air-core coil next to a aluminum plate that contains a right-cylindrical flat-bottom hole on the side opposite the coil. The half space below the coil can be considered as a two layer structure, where the first layer is the aluminum plate and the second layer is vacuum. A schematic drawing of the test is shown in Fig. 5.4. In the experiment, the impedance change of a precision wound coil of copper wire was measured at 20 frequencies, equally-spaced between 2.5 kHz and 50 kHz, with an HP 4194A impedance analyzer. Scans were repeated five times and the results averaged to produce the results reported herein. Measurements of the coil impedance were made as a function of frequency and position with respect to the center of the flat-bottom hole.

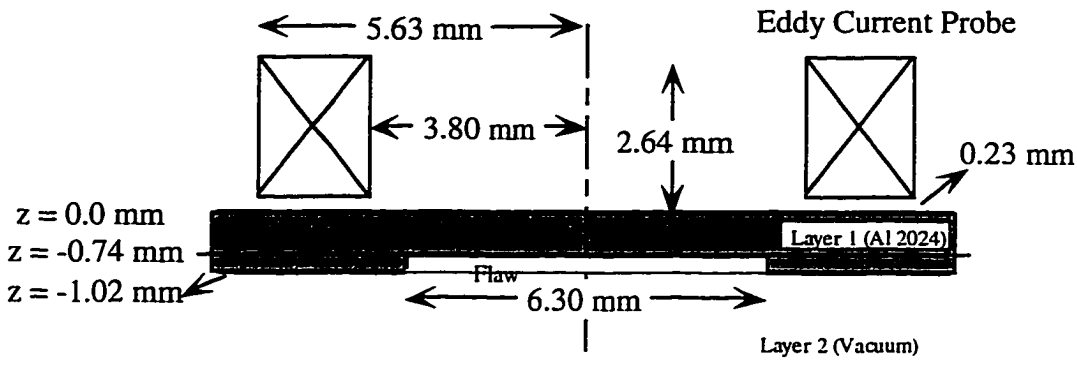


Fig. 5.4. Schematic drawing of the eddy current test problem of a flat bottom hole on the bottom side of a thin aluminum plate.

The coil has a right-cylindrical cross-section and consisted of 504 turns of copper wire wound on an insulating nonmagnetic core. The inner radius of the coil is 3.8 mm, the outer radius is 5.63 mm, and the height of the coil is 2.42 mm. The liftoff is 0.23 mm.

The test sample consists of a thin flat plate of 2024 aluminum alloy. The dimensions of the plate are 75 mm x 125 mm x 1.02 mm. A nominally right-cylindrical hole was drilled on the side of the plate opposite the coil. The diameter of the hole is 6.30 mm and the depth ranged between 0.28 mm at the center to 0.27 mm at the edge. This flaw is relatively large compared with the skin depth in the test frequency range and direct modeling by the volume integral method is not feasible due to the large computational resource requirement.

Implementation of the wavelet method

The modeling problem described above can be considered as the problem of determining the impedance change of the coil caused by a flaw in an infinite flat metal plate because the aluminum plate is very thin and the flaw is not close to the edges of the plate. Therefore, the problem is modeled using the thin plate geometry. The fast wavelet transform

algorithm used in our implementation is Mallat's pyramid algorithm for an orthonormal wavelet basis. It is worth mentioning that all the electric field quantities in the implementation are vectors with complex values. The wavelet transform of a complex function is obtained by transforming the real and imaginary parts independently and then combining the results. Also, to obtain $\mathbf{b}_{lmn}$ and $\overline{\mathbf{C}}_{ijklmn}$ in equation (5.30), multi-dimensional wavelet transforms must be used. Based on the approach we used to define the three-dimensional wavelets in equation (5.28), the multi-dimensional wavelet transform can be obtained by transforming the data in each dimension sequentially, and the result is not dependent on the order of the dimensions used in the transforms. This is exactly the same as for the multi-dimensional Fourier transform.

There exists another problem with the storage of the system matrix in the implementation of the wavelet method. The system matrix is the six-dimensional wavelet transform of the product of the Green's function and the conductivity distribution. The problem is how to compute the wavelet transform of the system matrix without fully storing it in the computer memory since this matrix could be very large when the number of volume elements used in modeling is sufficient for an accurate solution. To solve this problem, we utilized a property of Mallat's pyramid algorithm for the fast wavelet transform [46]. Notice that in the pyramid algorithm, the computation at any resolution level is solely based on the result of the immediately higher resolution level. Thus, we are able to compute a wavelet transform of size N from the results of wavelet transforms with a smaller size M if both N and M are powers of 2. Based on this relationship, we can compute the wavelet transform of the whole system matrix in two steps. In the first step, we divide the system matrix into

many submatrices with identical sizes and compute the wavelet transform for each of them. The result is stored in a sparse matrix data structure. In the second step, the wavelet transform of the whole matrix is obtained by using the result of the first step. The result is then thresholded and stored back in the sparse matrix data structure.

The steps used to implement the wavelet method are summarized as follows:

1. Compute the incident field distribution by using Dodd and Deed's result [28]. See Chapter 4 for details.
2. Compute the integral of the Green's function in each volume element. The Green's function for thin plate geometry is calculated by using the result of Weaver [31]. For singular elements, both analytical and numerical integrations are used to stabilize the result. For nonsingular elements, only numerical integration is used. Details on computing the Green's function for half space and thin plate geometry can be found in Chapter 4.
3. Compute the wavelet transform of the system matrix by using the two-step algorithm mentioned above. Implementation details of the pyramid algorithm for computing wavelet transform can be found in Reference [37].
4. Invert the transformed system matrix by using the sparse matrix LU decomposition routine.
5. Compute the wavelet transform of the incident field.
6. Compute the wavelet coefficients of the total field by backsubstitution.
7. Take the inverse wavelet transform to obtain the total field.
8. Calculate the impedance change by using the reciprocity principle formula (3.13).

Comparison of the numerical result and the experimental result

Based on the procedure discussed above, two numerical tests were performed for the test case. In the first numerical test, the aim was to compare the result obtained from the wavelet method by using the Haar basis and the periodic wavelet basis constructed from the Daubechies compactly supported wavelet with $p = 2$, which we will refer to as DAUB4P in later discussions for simplicity. In the first test, the test frequency was selected as 10 kHz, which corresponds to a skin depth of 1.17 mm in 2024 aluminum. The solution region was a rectangle with a length and a width equal to the diameter of the cylindrical flaw and a height equal to the height of the flaw. The solution region was modeled by 512 rectangular elements with 16 elements along each of the 2 horizontal directions and 2 elements in the vertical direction. In the first comparison, the system matrices were obtained by using a threshold of 0.005, which is relatively small, to ensure that a good result would be obtained. The matrix map for the sparse matrix obtained by using the Haar basis is shown in Fig. 5.5. After thresholding, only 4.36% of the total number of the matrix elements were nonzero elements. The matrix map for the sparse matrix obtained by using DAUB4P is shown in Fig. 5.6. In this case, 5.05% of all the matrix elements were nonzero elements after thresholding. The matrix structures shown in Fig. 5.5 and Fig. 5.6 indicate some of the properties of the wavelet method. First, most nonzero elements are close to the diagonal of the matrix, which indicates that a large interaction exists between the basis functions with overlapped or adjacent supports. Second, many nonzero elements are located in the upper-left corner of the matrix where low resolution wavelets are located, and only a few are located in the bottom-right corner where the high

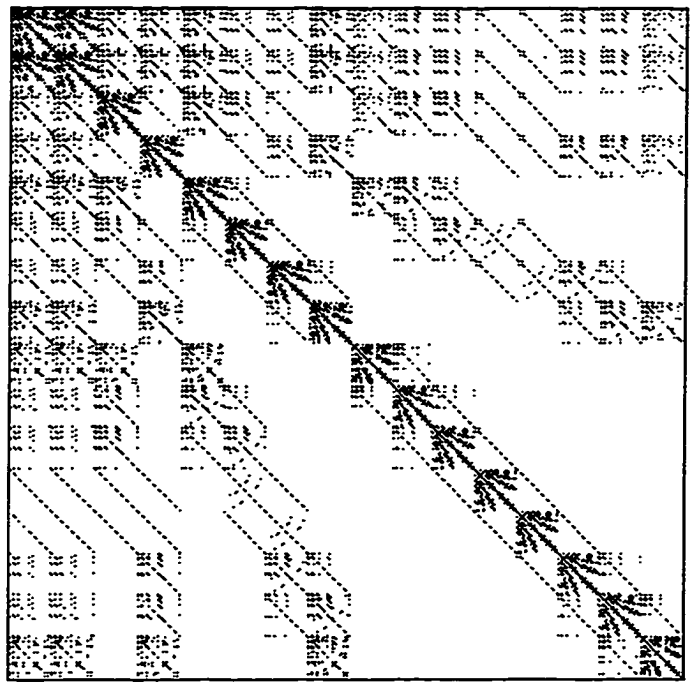


Fig. 5.5. The system matrix map after transformation by using the Haar wavelet and thresholded using a value of 0.005. Black spots denote nonzero elements.

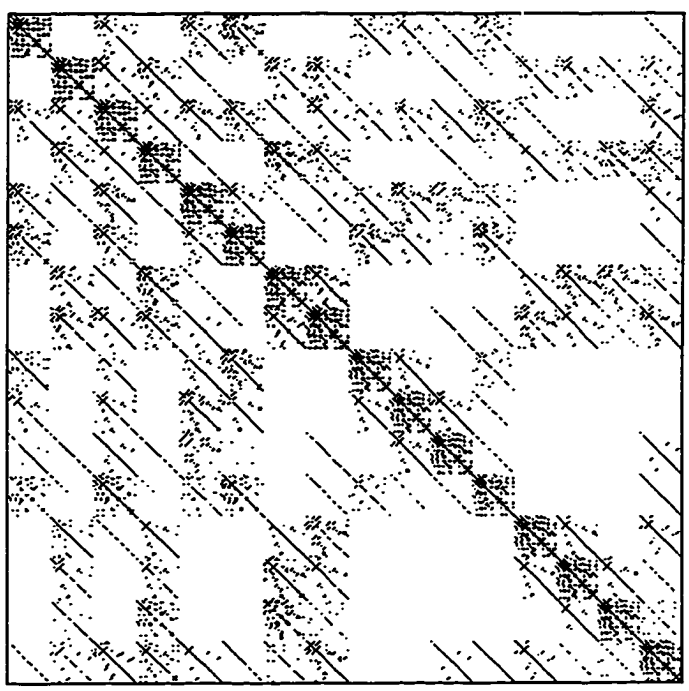


Fig. 5.6. The system matrix map after transformation by using DAUB4P and thresholded using a value of 0.005. Black spots denote nonzero element.

resolution wavelets are located. This implies that most interactions are between the low resolution components of the total field, and that the interactions between the two high resolution components that are not close to each other are very small. Third, the edge effect caused by periodization of the total field is obvious in Fig. 5.6, since there are many nonzero elements at the edges of the blocks at various resolutions. The impedance change results obtained by using the Haar basis and DAUB4P are compared with the experimental results in Fig. 5.7. The theoretical results are in good quantitative agreement with the experimental results. Also there is no significant difference between the result obtained using the Haar wavelet and the result obtained using the periodic wavelet in this case.

To examine the effect of the threshold on the accuracy of the results, or in other words, to decide whether there is some range of the threshold within which the results remain almost unchanged, we recomputed the results for the numerical example above using a threshold of 0.05. The matrix map for the sparse matrix obtained by using the Haar basis is shown in Fig. 5.8. After thresholding, only 0.413% of the matrix elements were nonzero. The matrix map for the sparse matrix obtained by using the DAUB4P is shown in Fig. 5.9. In Fig. 5.9, only 0.254% of the matrix elements were nonzero after thresholding. One can see that after using a threshold of 0.05, most nonzero elements are on the diagonal of the system matrix. The impedance change results in this case are compared with the experimental results in Fig. 5.10. Although the threshold is ten times larger and the thresholded matrices are rather sparse, an overall correspondence between the theoretical results and the experimental result can still be observed. It can be seen from this example that the wavelet method does indeed have a good potential for reducing the redundancy of the problem.

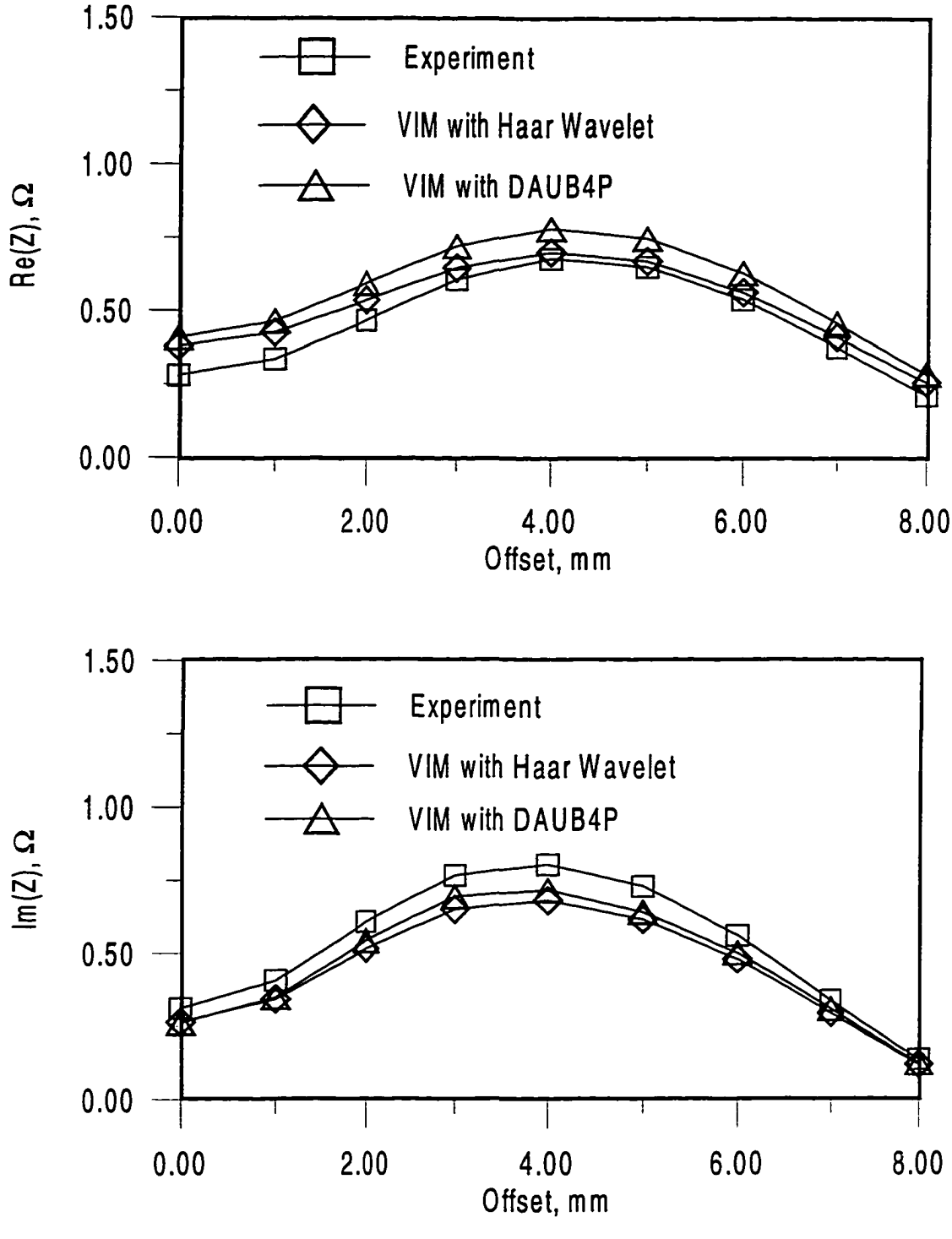


Fig. 5.7. Comparison of experimental results and the results obtained by using the Haar wavelet and DAUB4P with a threshold of 0.005.

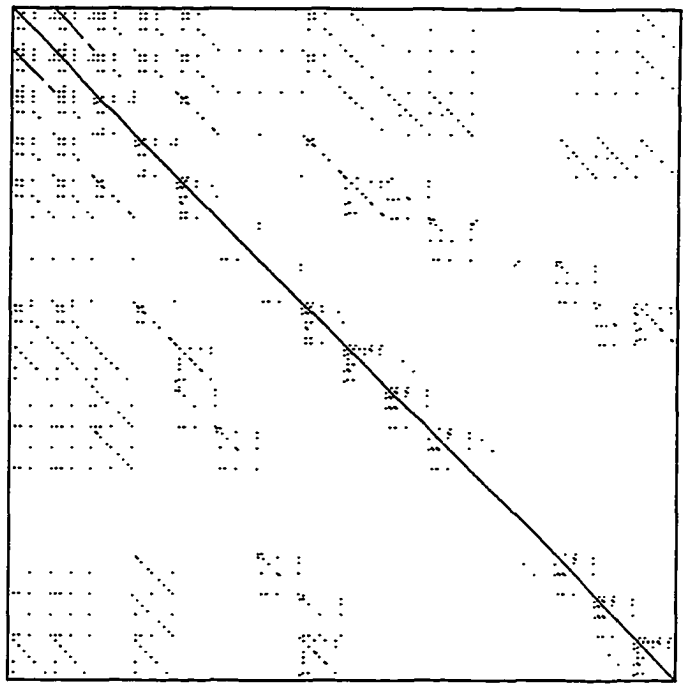


Fig. 5.8. The system matrix map after transformation by using the Haar wavelet and thresholded using a value of 0.05. Black spots denote nonzero elements.

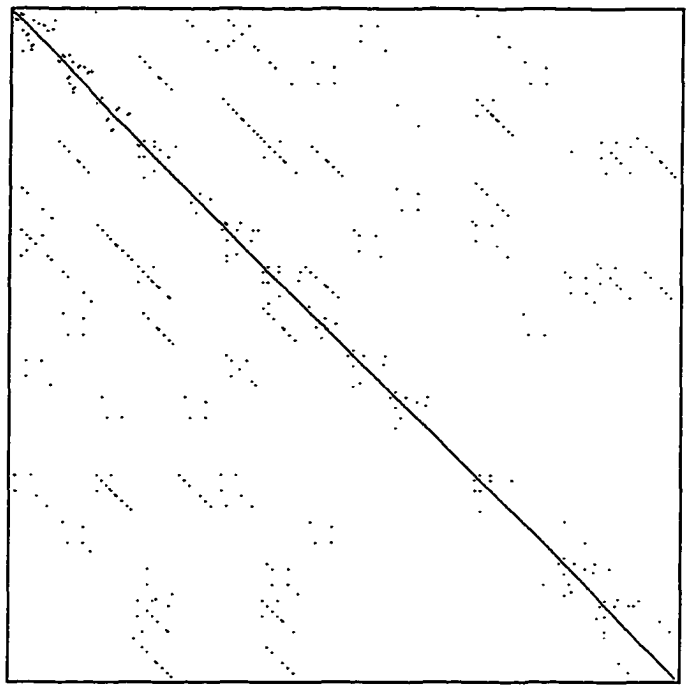


Fig. 5.9. The system matrix map after transformation by using DAUB4P and thresholded using a value of 0.05. Black spots denote nonzero elements.

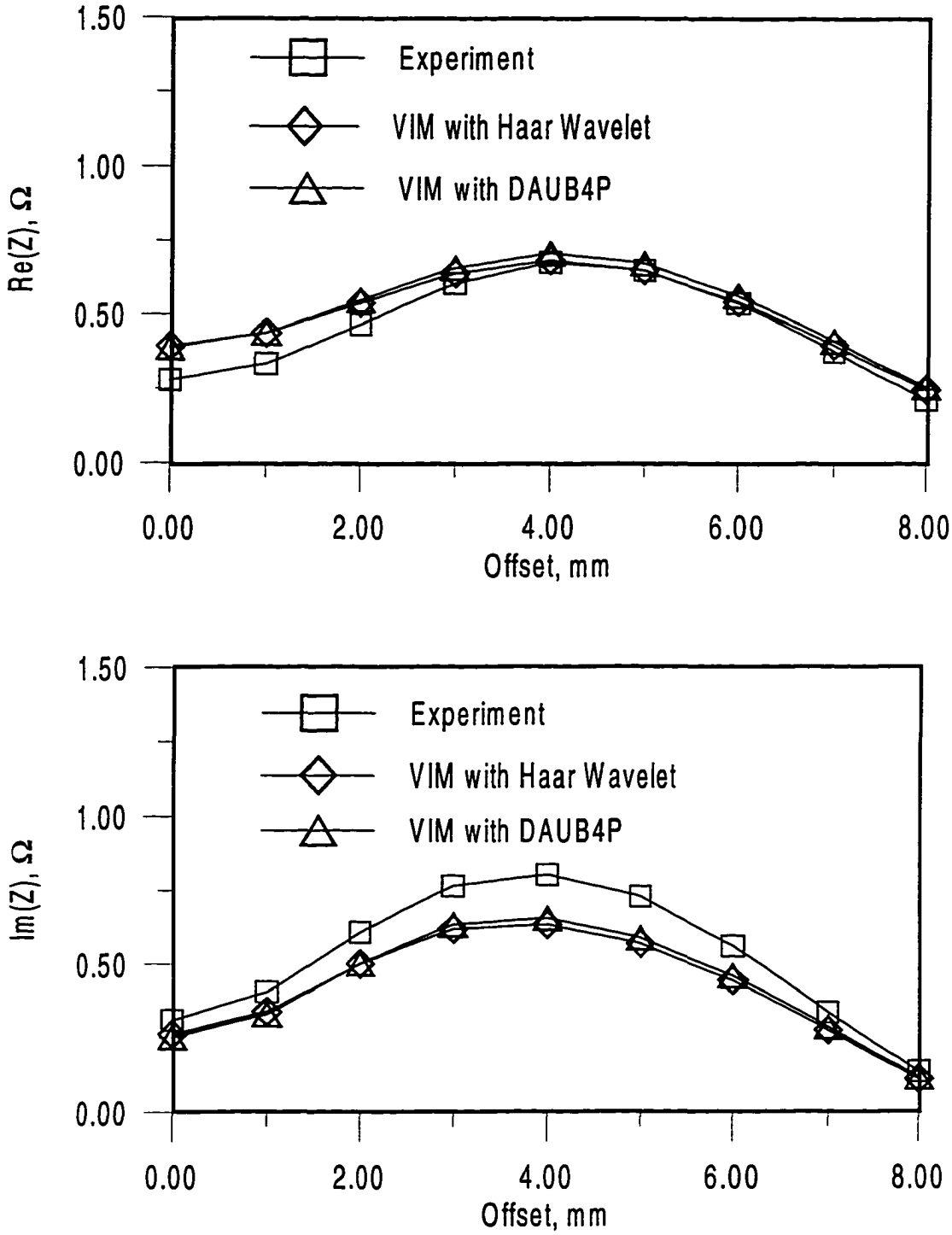


Fig. 5.10. Comparison of experimental results and the results obtained by using the Haar wavelet and DAUB4P with a threshold of 0.05.

The second numerical test was to examine the stability of the method over a relatively large frequency range. The theoretical impedance changes using DAUB4P and a threshold of 0.005 were computed for the frequency range from 2.5 kHz to 50 kHz, with a stepsize of 2.5 kHz, and for offsets from 0 mm to 8 mm, with a stepsize of 1 mm. The impedance change results as a function of frequency for offsets 0 mm, 4 mm and 8 mm are compared with the experimental results in Fig. 5.11. It is observed that the agreement between the results from the wavelet method and the experimental results are stable over this relatively large frequency range.

Discussions

In this chapter we have applied wavelet theory in the solution of the volume integral equation for the eddy current modeling problem. The new linear system equations after using the wavelet basis are obtained by applying a Galerkin method to the governing volume integral equation. The derivation is based on the result of the one-dimensional case, followed by the extension to the three-dimensional case and the extension to periodic wavelets. The wavelet method was tested using the problem of modeling hidden corrosion in a thin aluminum plate. Results obtained from both the Haar wavelet and the periodic wavelet constructed from the Daubechies compactly supported wavelet were compared with experimental results. The numerical examples show that by using the wavelet expansion method, it is possible to reduce the computational complexity by one to two orders with small loss in accuracy. This makes the eddy-current modeling problem of large three-dimensional flaws solvable using modest computing resources.

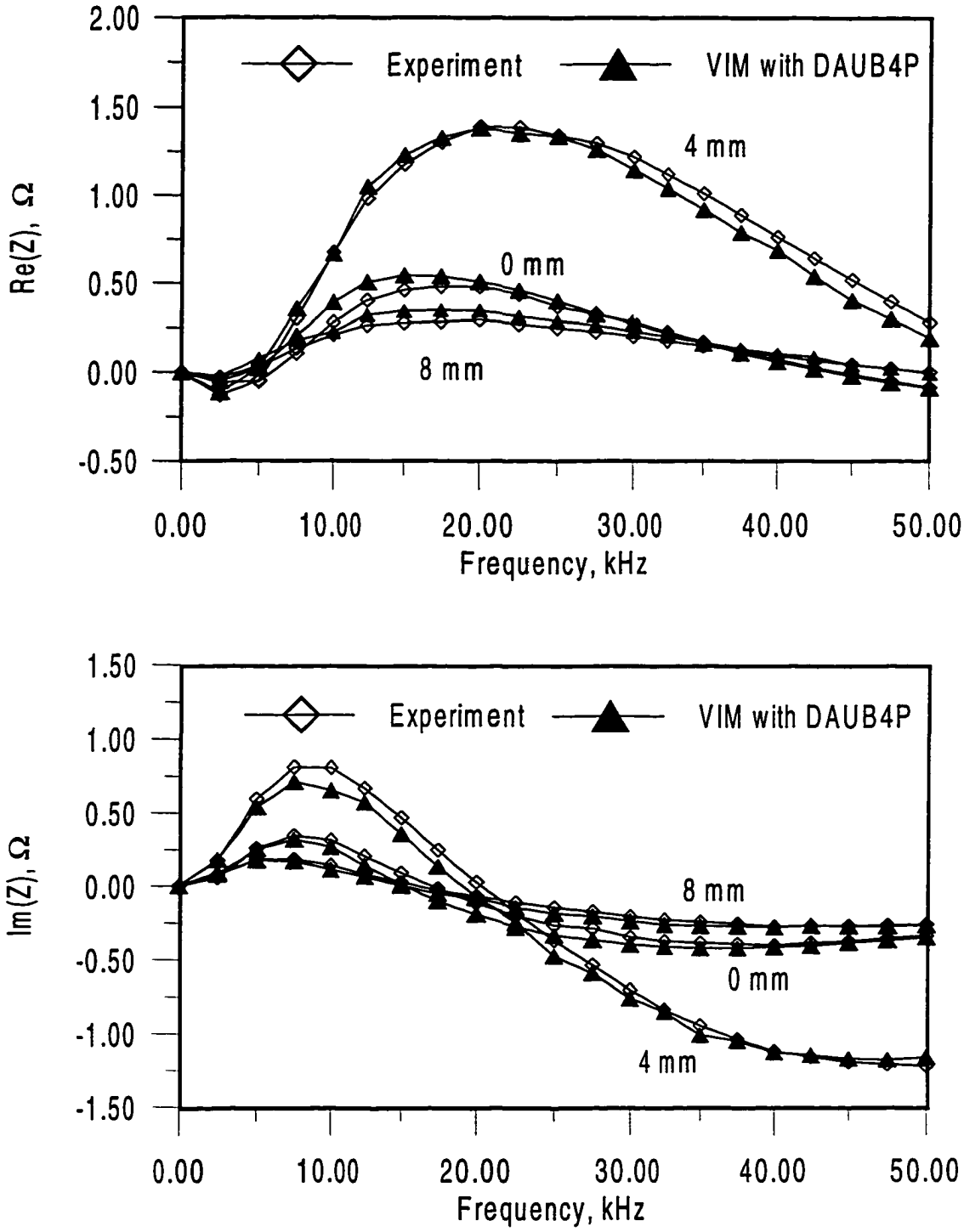


Fig. 5.11. Comparison of experimental results and the results obtained by using DAUB4P and a threshold of 0.005 for the frequency range from 2.5 kHz to 50 kHz.

CHAPTER 6. FAST EDDY CURRENT FORWARD MODELS USING ARTIFICIAL NEURAL NETWORKS

Introduction

In the preceding chapters we have discussed the volume integral method for eddy current forward modeling and a wavelet expansion method used to reduce the computational complexity of the volume integral method. Although in some cases the wavelet method can reduce the execution time by a factor of ten or more, the resultant execution time is still much longer than what is acceptable in some applications. One important application which belongs to this category is the solution of the nonlinear eddy current image restoration problem in which a large number of forward solutions must be computed. Another application is fast eddy current forward models for educational purpose in which the user would like to generate the modeling results for many forward problems in a short time.

In this chapter we apply artificial neural networks to the eddy current modeling problem in order to reduce the execution time of the forward model [50]. Our method is based on a two-dimensional imaging model in which an eddy current probe is considered as a black box transforming a flaw conductivity change image to a complex impedance change image. The nonlinear mapping from the flaw image to the impedance change image can be learned by using neural networks based on a training data set. After the learning process of the neural networks is finished, they can be used to generate outputs for new inputs.

There are two major reasons to use a neural network forward model. First, the computational complexity of most numerical models is $O(N^3)$, compared with $O(N^2)$ for

the forward computation of most neural networks, where N is the number of elements used in the forward model. This order of difference is significant when N is large. Second, the training data set for the neural network model can be obtained either theoretically or experimentally, which makes the neural network forward model capable of modeling complex geometry in which numerical models are hard to apply, but experimental measurements are still feasible.

The structure of the neural network eddy current forward model is shown in Fig. 6.1. The flaw image in Fig. 6.1 represents the two-dimensional conductivity change distribution of the flaw. If only cracks and voids are considered, a binary image can be used for the flaw image to reduce the complexity of the forward model. To reduce the size of the input image, i.e., to reduce the number of input features of the neural network, we use a two-dimensional Haar transform to capture the major characteristics of the flaw image. The inputs to the neural network are the thresholded Haar transform coefficients of the flaw image. The Haar

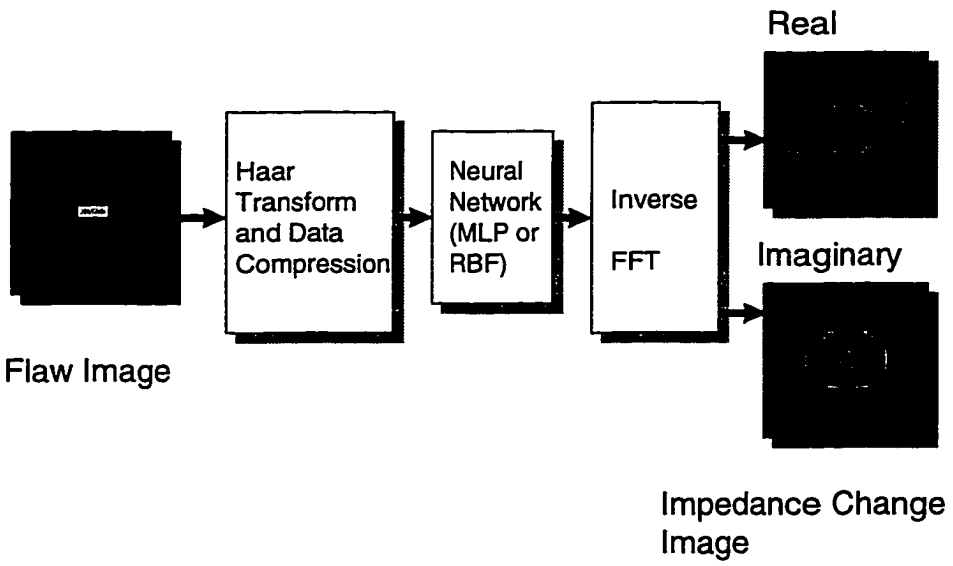


Fig. 6.1. Diagram of the neural network eddy current forward model.

transform is a wavelet transform with the mother wavelet being the Haar wavelet. The multiresolution decomposition capability of the Haar transform makes it easy to separate the major features of the flaw image from less important details of the flaw image. To reduce the dimensionality of the output space, we use the low frequency components of the impedance change image in the Fourier domain as the outputs of the neural network. The complex impedance change image is then obtained by applying the inverse FFT to the neural network outputs. The validity of this compression approach comes from the fact that the impedance change image is usually smooth due to the diffusive nature of eddy current.

Multilayer Perceptron

The multilayer perceptron (MLP) is one of the most widely used neural network models. A comprehensive discussion on the multilayer perceptron can be found in Reference [51]. Here we simply review some of its fundamental features. A multilayer perceptron has an input layer of sensory nodes (source nodes), one or more hidden layers of computational nodes, and an output layer of computational nodes. Nodes in adjacent layers are connected by synaptic weights. By changing the synaptic weights, we can change the functional form of the perceptron, and thus it can be used to approximate an unknown function. The multilayer perceptron is a uniform approximator, which means it can be used to approximate any smooth function to arbitrary accuracy if enough hidden layer nodes are used.

The forward computation of a multilayer perceptron is done on a layer-by-layer basis. First, for each hidden layer node or output layer node, an activation level is computed as

$$v_j^{(l)}(n) = \sum_{i=0}^{P} w_{ji}^{(l)}(n) y_i^{(l-1)}(n), \tag{6.1}$$

where $v_j^{(l)}(n)$ is the activation level of node j in layer l at the time instant n, $w_{ji}^{(l)}(n)$ is the synaptic weight between node j in layer l and node i in layer $l-1$ at time instant n, $y_i^{(l-1)}(n)$ is the output of node i in layer $l-1$ at time instant n, and P is the number of nodes in layer $l-1$. Apparently the activation level is the result of the innerproduct operation between the input vector (the vector containing all output values in layer $l-1$) and the weight vector of node j in layer l. The output of a hidden layer node or an output layer node is a nonlinear function of its activation level, which is usually called sigmoidal nonlinearity. A particular form of the sigmoidal function is the logistic function

$$y_j^{(l)}(n) = \frac{1}{1+\exp(-v_j^{(l)}(n))}, \tag{6.2}$$

which is a monotonic increasing function bounded in 0 to1.

The training of a multilayer perceptron is usually done by using the backpropagation learning algorithm. The backpropagation algorithm is a gradient based iterative algorithm in which the learning error is propagated backwards through the network. Accordingly, the synaptic weights of the output layer are updated first, and then the synaptic weights of the hidden layer next to the output layer, and so on. The synaptic weight update equation in the backpropagation algorithm is given by

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n), \tag{6.3}$$

where η is a learning rate parameter, and $\delta_j^{(l)}(n)$ is the local gradient for node j in layer l. For an output layer node, the local gradient is given by

$$\delta_j^{(L)}(n) = [d_j(n) - o_j(n)]o_j(n)[1 - o_j(n)], \tag{6.4}$$

where $d_j(n)$ is the desired output of output node j, and $o_j(n)$ is the actual output of output node j, i.e., $o_j(n) = y_j^{(L)}(n)$. For a hidden layer node, the local gradient is given by

$$\delta_j^{(l)}(n) = y_j^{(l)}(n)[1 - y_j^{(l)}(n)]\sum_k \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n). \tag{6.5}$$

Equation (6.5) indicates that the local gradient for node j in layer l is related to the local gradients and synaptic weights for all nodes in layer $l+1$. Therefore, it can be computed only after the local gradients in the next layer have been computed. Consequently, the computation of the backpropagation algorithm is also on a layer-by-layer basis, starting from the output layer.

One major disadvantage of the backpropagation algorithm is that it is slow in convergence, especially for large-scale problems. The slow convergence is a result of using the gradient based weight update formula since gradient based methods tend to stick to local minima [52]. To improve the convergence performance of the original backpropagation algorithm, we have used several modifications such as adding a momentum term, using an asymmetric sigmoidal function like hyperbolic tangent instead of a logistic function, and applied the Delta-Bar-Delta learning rule [51] to adaptively control the learning rate for faster convergence. Our experience has shown that these modifications, especially the Delta-Bar-Delta learning rule, can significantly improve the convergence speed of backpropagation. However, due to its inherent local optimization property, the learning of the backpropagation algorithm is still too slow for problems like the eddy current forward modeling. Our

experience has shown that the learning time becomes impractical even when the training data set only contains several examples.

Radial Basis Function Neural Networks

The radial basis function (RBF) neural network is another important class of feedforward layered neural networks. Theories and applications of the RBF networks can be found in References [53-60]. An RBF network has one input layer, one hidden layer and one output layer. The nonlinearity of an RBF network is implemented in the hidden layer. The output layer of an RBF network is a linear layer, compared with the nonlinear output layer in a multilayer perceptron. The major application of an RBF network is functional approximation [57], i.e., to approximate an unknown function given the values of the function on some sample data points. This approximation problem is ill-posed because much of the information of the function is not available so there are many possible solutions. To deal with the ill-posedness of the approximation problem, regularization schemes must be applied to incorporate a priori information in the solution. The RBF network can be thought of as one solution of this regularization problem. It is also a uniform approximator.

The output of a hidden layer node in an RBF network is given by

$$\phi_i(\mathbf{X}) = g\left(\left\|\mathbf{X} - \mathbf{C}_i\right\|\right), \tag{6.6}$$

where $\mathbf{X}$ is the input vector, $g(\|\cdot\cdot\|)$ is a radial basis function, and $\mathbf{C}_i$ is the center vector of the radial basis function. Usually the norm used in equation (6.6) is the 2-norm, or the Euclidean distance. There are many possible forms for the radial basis function g. However, the most widely used is the multivariate Gaussian

$$G_i(\mathbf{X}) = \exp\left[-\frac{1}{2}(\mathbf{X}-\mathbf{C_i})\Sigma_i^{-1}(\mathbf{X}-\mathbf{C_i})\right], \tag{6.7}$$

where $\mathbf{C_i}$ is the center of the Gaussian function, and Σ_i is the covariance matrix of the multivariate Gaussian function. It is clear that $\mathbf{C_i}$ determines the location of the Gaussian function and Σ_i determines the shape of the Gaussian function.

As mentioned above, the output y_j of node j in the output layer is a linear combination of the outputs in the hidden layer

$$y_j = \sum_i w_{ji} G_i(\mathbf{X}), \tag{6.8}$$

where w_{ji} is the linear weight connecting node j in the output layer and node i in the hidden layer.

The training of the RBF network finds the optimal values for $\mathbf{C_i}$, Σ_i, w_{ij} so as to minimize the learning error. In general, a nonlinear optimization algorithm must be used for the training of the hidden layer parameters, and a linear optimization algorithm must be used for the output layer parameters. In the literature, there are various learning algorithms [51, 58, 61, 62] proposed for the RBF network in different applications. For our particular problem, we considered two cases:

Case 1 - small number of training samples

In this case, we can simply use the training sample inputs as the centers for the radial basis functions, i.e.,

$$\mathbf{C_i} = \mathbf{X_i}, \quad i = 1, \cdots, M, \tag{6.9}$$

where $\mathbf{X_i}$ are the input vectors of the training samples, and M is the number of training samples. The shape of the multivariate Gaussian function is decided by a diagonal covariance matrix

$$\Sigma = diag\left\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2\right\}, \tag{6.10}$$

where N is the number of inputs, $\sigma_i = \frac{d}{\sqrt{2M}}$, $i = 1, \cdots, N$, and d is the maximum distance between the centers of the Gaussian functions. In this case, the equal potential surface of each multivariate Gaussian function is a hypersphere. To decide the optimal output layer weights, we need to solve

$$\mathbf{Y} = \mathbf{GW}, \tag{6.11}$$

where $\mathbf{Y}$ is a matrix containing the desired outputs given by the training samples, $\mathbf{G}$ is a matrix containing hidden layer outputs, and $\mathbf{W}$ is the matrix containing output layer weights. Since in this case the number of hidden layer nodes is equal to the number of training samples, equation (6.11) is well-defined (equal number of equations and unknowns) and can be solved directly by

$$\mathbf{W} = \mathbf{G}^{-1}\mathbf{Y}, \tag{6.12}$$

where the matrix inversion can be done by using LU decomposition and back substitution.

Case 2 - large number of training samples

In this case, we cannot use training sample inputs as centers for RBFs because this will lead to a network with a large number of hidden layer nodes which is slow and difficult to train. To reduce the number of centers, we use the K-Mean algorithm [51] to find K

cluster centers in M training samples with K much smaller than M. For the K multivariate Gaussian functions, the diagonal covariance of equation (6.10) can still be used, however, now we let $\sigma_i = \frac{d_i}{\sqrt{2M}}$, $i = 1, \cdots, N$, and

$$d_i = \max_{j,k} \left\{ \left| C_j^i - C_k^i \right| \right\} \tag{6.13}$$

is the maximum distance between the RBF centers in the i th dimension. As a result of equation (6.13), the equal potential surface of each multivariate Gaussian function becomes a hyperellipse. In this case, the optimal output layer weights cannot be computed directly, because now there are more equations than unknowns in equation (6.11) so that the problem is overdetermined. We can use a Least Mean Square (LMS) algorithm to iteratively compute the optimal output layer weights [51]

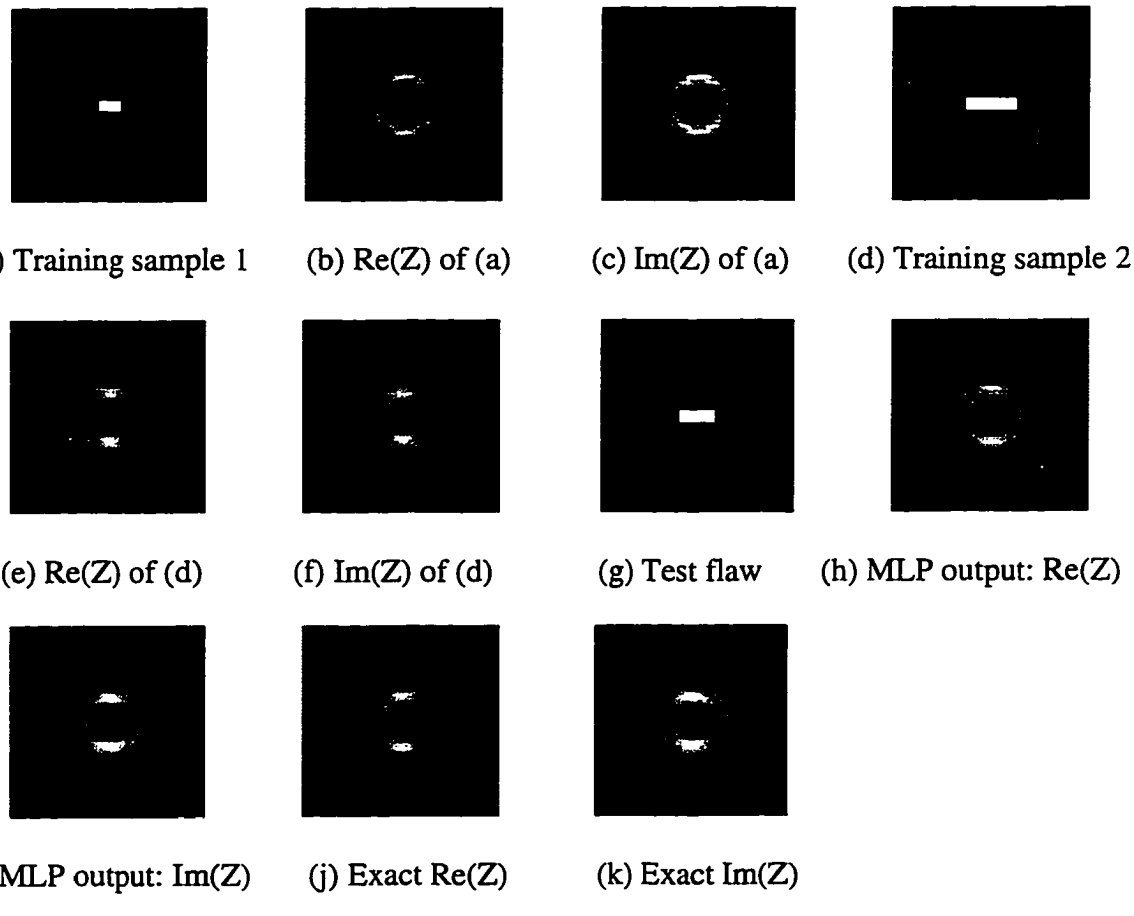
$$w_{ji}(n+1) = w_{ji}(n) + \eta(y_j - o_j)g_i, \tag{6.14}$$

where w_{ji} is the weight between output layer node j and hidden layer node i, y_j is the desired output of output layer node j, o_j is the actual output of the output layer node j, g_i is the output of the hidden layer node i, and η is a learning rate parameter. A disadvantage of equation (6.14) is that it is much slower than the direct matrix inversion in equation (6.12). Therefore, in practice equation (6.12), the Case 1 algorithm, is more frequently used.

Numerical Results

To illustrate some properties of the neural network forward model, we numerically tested both the MLP and the RBF models. The first test was to examine the generalization

property of MLP. We used a short crack and a long crack, and the corresponding impedance change images as a training data set to train a three layer perceptron. After the learning process was finished, a crack with medium length was used to test the perceptron. The outputs of the neural network model were then compared with exact solutions obtained by using a volume integral model with wavelet expansion. The training data set and the test results for the first test are shown in Fig. 6.2. As can be seen in Fig. 6.2, the impedance change images given by the neural network model are rather close to the exact results, although some minor differences are still visible.



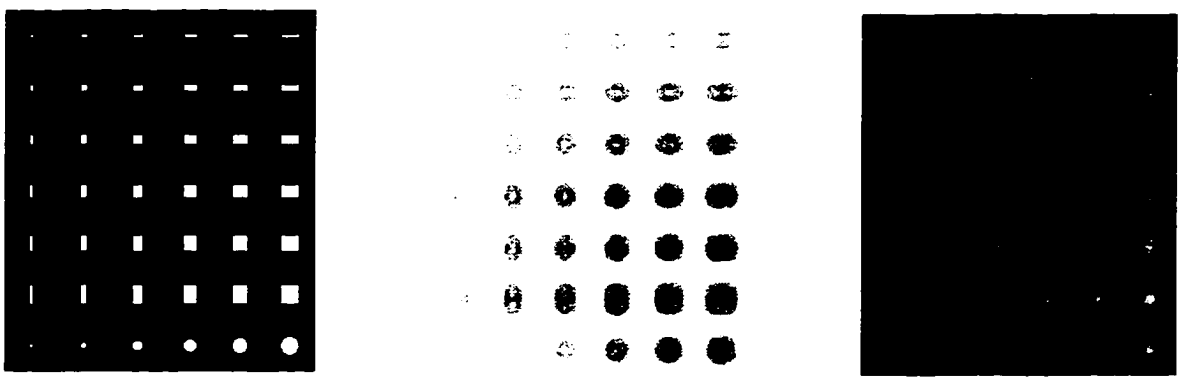
(a) Training sample 1 (b) Re(Z) of (a) (c) Im(Z) of (a) (d) Training sample 2

(e) Re(Z) of (d) (f) Im(Z) of (d) (g) Test flaw (h) MLP output: Re(Z)

(i) MLP output: Im(Z) (j) Exact Re(Z) (k) Exact Im(Z)

Fig. 6.2. Test results on the generalization capability of multilayer perceptron.

Although the test on the generalization property of MLP is rather satisfying, our experience indicates that it is not a practical method to our particular problem due to its extreme slow convergence speed. It is only useful when the number of training samples in the training data set is very small. On the other hand, our experience has shown that the learning process of the RBF network is faster even for a relatively large training data set. Fig. 6.3 gives a training example of the neural network forward model based on an RBF net. In this training process, we used 36 slots with various length and width and 6 holes with various radius as training flaws. To find the impedance change images for the 42 flaws, we



(a) Training flaw images (b) Re(Z) images (c) Im(Z) images

Fig. 6.3. Training example of the radial basis function neural network.

used the volume integral method code. The computation of the 42 forward problems took about 10 hours of CPU time on a DEC 5000 workstation. Then we trained an RBF network using these 42 training samples. The learning process took about 45 seconds using the Case 1 algorithm and it only took 30 seconds for the RBF forward model to regenerate the 42 training samples. Compared with the 10 hours used by the numerical model, this is over one

thousand times of speed improvement. To test the generalization capability of this RBF forward model, we used an elliptical test flaw which was not used in the training data set. The output of the RBF forward model and the exact outputs are compared in Fig. 6.4. As can be seen in Fig. 6.4, it is very difficult visually to tell the difference between the RBF outputs and the exact outputs from the images. Therefore, we made some quantitative comparisons. The normalized maximum error between the two results is 1.6%, and the normalized mean square error between the two results is only 0.32%. We judge this accuracy to be acceptable for most real world applications.

(a) Test flaw (b) RBF Re(Z) (c) RBF Im(Z) (d) Exact Re(Z) (e) Exact Im(Z)

Fig. 6.4. Test results of radial basis function neural network.

Conclusions

In this chapter we discussed how to use artificial neural networks to build fast eddy current forward models. Our numerical results showed that these models can give several orders of speed improvement over traditional numerical models. We found that one important step in applying this method is to create a proper training data set. For the two neural network models we studied, the RBF net is considered more practical due to its faster learning process, although the generalization property of MLP may be better.

CHAPTER 7. PREPROCESSING TECHNIQUES FOR TWO-DIMENSIONAL EDDY CURRENT INSPECTION DATA

Introduction

Inspection data obtained from two-dimensional eddy current inspection using a MRPC are usually distorted by noise and many other unwanted effects. Noise in the inspection data include electromagnetic interference from the driving motor of the MRPC probe, and quantization noise and thermal noise in the data acquisition circuit. Changes in probe liftoff, tube diameter, and tube wall thickness also create large background signals which make recognition and characterization of flaw signals difficult. Furthermore, due to variations in probe characteristics and instrumentation setup, a calibration procedure must be performed to align the phase and magnitude of the inspection data with a calibration standard. This calibration step is usually a time-consuming process. In this chapter we discuss several preprocessing techniques which can be used to remove some unwanted signals in the data and to prepare the data for further signal and image processing procedures. The preprocessing techniques we have studied include background removal methods to remove liftoff signals and an automatic phase adjustment algorithm which can be used to simplify the calibration process in some situations. The preprocessing techniques discussed in this chapter are used only as the first step of the processing. Therefore, further processing must be performed to obtain good data analysis results. Due to space limit, processing results are not included in this Chapter. They can be found in Reference [63].

Background Removal

The first step of preprocessing is to remove the background signal in the raw data. Due to the diffusive nature of eddy current testing, the coil impedance change is very sensitive to liftoff and small variations on tube internal surface. Because the background signal is generally much larger than the flaw signal (especially for a MRPC probe), it must be removed from the raw data before any phase rotation can be applied to the data. We have tested three methods for background removal and they are described in this section.

Background removal by polynomial fitting

The first method uses a least-square polynomial fitting to obtain the background signal and then the background signal is subtracted from the original data [64]. The background removal of the horizontal and the vertical components of the raw data are performed independently. Also, because the data is two-dimensional, the background removal is first performed on the horizontal direction, and then on the vertical direction. Given a data set $f(x_i)$, $i = 1, \cdots, n$, the problem of fitting a mth order polynomial

$$g(x) = \sum_{j=0}^{m} a_j x^j \tag{7.1}$$

is to find the group of coefficients $\{a_j\}$ so that the least-square error

$$\varepsilon = \sum_{i=1}^{n} \left[f(x_i) - \sum_{j=0}^{m} a_j x_i^j \right]^2 \tag{7.2}$$

is minimized. Setting

$$\frac{\partial \varepsilon}{\partial a_l} = 0, \quad l = 0, \cdots, m, \tag{7.3}$$

we have the following $m+1$ equations for $m+1$ unknowns

$$\sum_{j=0}^{m} a_j \left[\sum_{i=1}^{n} x_i^{j+l} \right] = \sum_{i=1}^{n} f(x_i) x_i^l, \quad l = 0, \cdots, m. \tag{7.4}$$

These linear equations can be solved by using Gauss-Jordon elimination and backsubstitution.

Normally a second-order polynomial fitting is used for both horizontal and vertical directions. Higher order fitting generally is not necessary and may introduce artifacts in the result. Even with the second-order polynomial fitting, when there is a strong signal in the data, the background obtained is usually distorted by the presence of the large signal. This will generate some artifacts near the large signal in the background removal result.

Background removal by median filtering

The second method is to smooth the data by using median filtering [65] and then subtract the smoothed version from the original data. A nice property of the median filter is that it can remove spike-like noise without removing edges in the data. Therefore, the smoothed signal will not be distorted by a large, pulse-like signal if a median filter with the proper window size is used. Because our objective is to detect circumferential cracks, anything wider than the typical width of a crack signal (which can be decided by the characteristics of the probe because the crack itself is usually very narrow) can be considered as part of the background signal. Thus, we can use a median filter along the vertical direction with a window length larger than two times the typical width of a crack signal to obtain the

background. The disadvantages of median filtering are that it is relatively slow because at each pixel in the image a sorting of local data values must be performed, and it has an edge effect because a running window is used.

Background removal by masked polynomial fitting

Since both the polynomial fitting method and the median filtering method have disadvantages, we investigated a third method which somewhat combines the first two methods. This method is a modification of the polynomial fitting method, but it also uses the idea of order statistics. In this method, before the polynomial fitting, the data is sorted, and the parts of the data with the largest values and smallest values will not be used in the polynomial fitting. For example, if given a margin of 20%, the data points with values in the range of the largest 20% and the smallest 20% will not be used in polynomial fitting after sorting. Therefore, if there is a strong but localized signal, the data points from this signal will not be used in the polynomial fitting which results in fewer artifacts in the background removal result. Also, this method is almost as fast as the direct polynomial fitting method because in this case only one sorting operation is required for each line of data. We do not need to perform a sorting operation for each pixel.

Automatic Phase Adjustment

The second step of the preprocessing is to adjust the phase of the raw data so that in one channel the flaw signal is maximized and all other unwanted signals are minimized. This step is necessary because in later processing we want to deal with a real image, and we want

the signal-to-noise ratio (SNR) of this data to be maximized. When the SNR of the raw data is high, it is not hard to adjust the phase manually since the flaw signal is visible in the raw data. However, if the SNR of the raw data is low, it is almost impossible to see the flaw signal in the raw data and the phase adjustment is difficult and time consuming without a standard calibration procedure. To deal with this problem, we developed an automatic phase adjustment method which can find the optimal rotation angle without performing the calibration procedure. This method is based on the assumptions that there exists a phase difference between the signal and the noise, and the energy of the noise is much larger than that of the signal. The first assumption is valid because we are interested in OD cracks, and from eddy current theory they have a phase difference dependence upon the liftoff signal and ID noise. The second assumption simply means low SNR.

To explain the method, let us consider a simple noisy signal which is the sum of two vectors, the signal vector and the noise vector, respectively

$$A_s e^{j\omega\theta_s} + A_n e^{j\omega\theta_n}$$

where $A_n >> A_s$ and there is a nontrivial difference between θ_s and θ_n. Now if we rotate this signal to make the SNR maximized for the vertical (or imaginary) component, then the best rotation angle is the one that rotates the noise vector exactly to the horizontal axis so there is no projection of the noise vector on the vertical axis. It can be proved in this case that the energy of the horizontal component is also maximized. From this simple example, we see that the problem of finding the rotation angle that gives the largest SNR in the vertical channel is equivalent to the problem of finding the rotation angle for maximizing the energy in the horizontal channel.

Without loss of generality, consider a one-dimensional data set

$$\{X_i = R_i + jI_i\}, \tag{7.5}$$

where the energy of the horizontal component of the data after rotating by a phase angle θ can be expressed as

$$E = \sum_i \left[\mathrm{Re}\{X_i e^{j\theta}\}\right]^2. \tag{7.6}$$

To find the θ for maximizing E, set

$$\frac{\partial E}{\partial \theta} = 0. \tag{7.7}$$

Because

$$\mathrm{Re}\{X_i e^{j\theta}\} = R_i \cos\theta - I_i \sin\theta, \tag{7.8}$$

we have the following equation after differentiation,

$$\sum_i (R_i \cos\theta - I_i \sin\theta)(-R_i \sin\theta - I_i \cos\theta) = 0. \tag{7.9}$$

After some manipulation, the above can be written as

$$\sum_i \left[\frac{1}{2}(R_i^2 - I_i^2)\sin 2\theta + R_i I_i \cos 2\theta\right] = 0. \tag{7.10}$$

Then it follows that

$$\theta = \frac{1}{2}\arctan\frac{\sum_i R_i I_i}{\sum_i (I_i^2 - R_i^2)}. \tag{7.11}$$

The above derivation can be easily extended to a two-dimensional case. The above equation can still be used, except now the summations are on all the data points in the image

$$\theta = \frac{1}{2}\arctan\frac{\sum_i \sum_j R_{ij} I_{ij}}{\sum_i \sum_j (I_{ij}^2 - R_{ij}^2)}. \tag{7.12}$$

Test results [63] have shown that when the assumptions of the automatic phase rotation algorithm are satisfied, it can provide a result similar to that of the standard calibration procedure. However, when the data has relatively high SNR or when there are signals from OD tube variations which have very similar phase as the OD crack signal, the standard calibration process is still necessary.

CHAPTER 8. EDDY CURRENT IMAGE PROCESSING USING PRINCIPAL COMPONENT ANALYSIS

Introduction

In this chapter we present an eddy current image processing technique based on principal component analysis (PCA) [66-68]. PCA is a widely-used signal processing and image compression technique which has many applications in feature extraction, data compression, and noise reduction [51]. In our problem, it is used as a noise reduction method to remove the background signal in the two-dimensional eddy current inspection data. In eddy current measurements, liftoff and other ID variations create large background signals, which usually make small crack signals hard to be recognized. The background removal methods discussed in the last chapter can be used to remove global background signals in the data. However, for background signals that have similar spatial distributions as the flaw signals, these preprocessing techniques are very inefficient. The background signals with smaller spatial supports usually result from dents and other ID variations that are common for a heavily used tube. In order to remove these background signals, we developed the background removal method based on PCA. In our method, PCA is used to estimate the background of the image and then to subtract the estimate from the original image. We have finished some tests on the PCA method using MRPC data. The test results have shown the high potential of the method. It is hoped that this technique will be helpful to human analysts in detecting small crack signals.

Basic Theory of PCA

What is PCA?

Principal Component Analysis [66-68] is a statistical method which determines an optimal linear transformation

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \tag{8.1}$$

to transform an input vector $\mathbf{X} \in R^n$ of a zero-mean, wide-sense stationary stochastic process to a vector $\mathbf{Y} \in R^m, m < n$ so that the components of $\mathbf{Y}$ are uncorrelated. PCA is also referred to as the Karhunen-Loeve or Hotelling transform. PCA has many important applications, such as spectral analysis, pattern recognition and image compression. It can be proven that the optimal transform matrix $\mathbf{W}$ can be constructed from the first m orthonormal eigenvectors $\{\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_m\}$ of the covariance matrix of $\mathbf{X}$

$$\mathbf{R}_{\mathbf{XX}} = E\{\mathbf{X}\mathbf{X}^{\mathbf{T}}\}. \tag{8.2}$$

The eigenvectors are normalized so that

$$\mathbf{W}_i^{\mathbf{T}}\mathbf{W}_j = \delta_{ij}. \tag{8.3}$$

The eigenvectors and eigenvalues are solutions to the following eigenvalue problem

$$\mathbf{R}_{\mathbf{XX}}\mathbf{W}_j = \lambda_j \mathbf{W}_j, \tag{8.4}$$

and the first m eigenvectors correspond to the m largest eigenvalues. The first m eigenvectors are considered the principal eigenvectors, which are directions in the n-dimensional vector space in which the input data have the largest variance. By projecting the input data through $\mathbf{W}$, the n-dimensional input vector is mapped into the m-dimensional subspace spanned by the principal eigenvectors. In other words, the input data is

compressed, which means PCA can transform a large amount of correlated input data into a set of statistically decorrelated components.

PCA using unsupervised Hebbian learning

The application of the PCA method discussed above requires the solution of the eigenvalue problem for the covariance matrix of the input data. The direct solution of eigenvalue problems by matrix decomposition is complex and time-consuming. Therefore, a neural network approach, the Hebbian learning network [51], has been developed to solve the eigenvalue problem. Another advantage of the Hebbian learning neural network is that it does not require the computation of the covariance matrix, but directly obtains the principal eigenvectors from the input data.

A Hebbian learning net is a simple two layer linear network where the first layer contains the elements of the input vector, and the second layer contains the elements of the output vector. An output of the network is given by

$$y_j = \sum_i W_{ij} x_i, \tag{8.5}$$

where W_{ij} are the synaptic weights. Initially the weights of the network are set to random numbers. The network is then trained by feeding input data into it and the weights are iteratively modified. The objective is to force the weights to finally converge to the principal eigenvectors of the covariance matrix of the input data. By assuming that the input data is ergodic, the covariance matrix can be expressed as

$$\mathbf{R}_{XX} = \sum_i \mathbf{X}_i \mathbf{X}_i^T, \tag{8.6}$$

where the summation is taken over all the input data. To find the weight update equation which forces the weights to converge to the principal eigenvectors, we need to minimize the energy

$$E(\mathbf{W_j}) = \frac{1}{2}\left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_2^2 \tag{8.7}$$

under the normalization condition

$$\mathbf{W_j}^{\mathrm{T}}\mathbf{W_j} = 1, \tag{8.8}$$

where $\hat{\mathbf{X}} = \mathbf{W_j} y_j$ under the normalization condition. Notice that in the case of a single output node, the energy is minimized if the weight vector converges to the largest eigenvector of the covariance matrix of the input data. Minimization of the energy can be achieved by using a gradient descent optimization method. Taking the partial derivative of the energy, we have

$$\frac{\partial E(\mathbf{W_j})}{\partial \mathbf{W_j}} = -y_j(\mathbf{X} - \mathbf{W_j} y_j), \tag{8.9}$$

which is the direction in which the energy has the fastest increase. We can reduce the energy by updating the weight vector along the opposite direction of the gradient, e.g.,

$$W_{i1}(k+1) = W_{i1}(k) + \mu(k) y_1(k)(x_i(k) - W_{i1}(k) y_1(k)), \tag{8.10}$$

for the case of a single output node, where $\mu(k)$ is the learning rate. The above weight update equation is called Oja's learning rule [66]. It can be extended to the case of m output nodes (m principal eigenvectors) as

$$W_{ij}(k+1) = W_{ij}(k) + \mu(k) y_j(k)[x_i(k) - \sum_{h=1}^{m} W_{ih}(k) y_h(k)]. \tag{8.11}$$

However, Oja's learning rule as given above does not converge to the first m principal eigenvectors of $\mathbf{R}_{XX}$, but rather to a non-unique linear combination of the first m eigenvectors of $\mathbf{R}_{XX}$ which still span the subspace of the m principal eigenvectors of $\mathbf{R}_{XX}$. There is a generalized Hebbian algorithm [68] which converges to exactly the first m principal eigenvectors of $\mathbf{R}_{XX}$ in descending eigenvalue order. The only difference in the generalized learning rule is the upper limit in the summation, i.e.,

$$W_{ij}(k+1) = W_{ij}(k) + \mu(k) y_j(k)[x_i(k) - \sum_{h=1}^{j} W_{ih}(k) y_h(k)]. \qquad (8.12)$$

Eddy Current Background Removal Using PCA

Basic idea

As mentioned above, PCA has been widely used in data compression, especially image compression applications. The power of the PCA method in data compression applications is its capability to decorrelate a large amount of data and find the principal components (or modes) of the data. In other words, by using PCA, we can reconstruct the image with only the expansion coefficients for several principal eigenvectors and the reconstruction will still be close to the original data. What will be lost is the detail of the image, which has a spatial variation different from the major trend (the principal components, or modes) of the image and corresponds to smaller eigenvalues. Because the eigenvalues corresponding to the detail are much smaller than the principal eigenvalues, they are associated with only a small portion of the total energy of the image. Therefore, even though

this part of the data is not recovered in the reconstruction results, the mean square error of the reconstruction is still very small.

But how does this relate to the processing of eddy current images? Recall that a major problem in eddy current signal processing is to remove the background signal caused by liftoff and ID variation. We can find a similarity between the problem of characterizing the background and the problem of compressing an image. The background signal mentioned here is not referred to as the overall background of the data, but rather the part of the background that has faster spatial variation and a more complex spatial pattern that may be or may not be periodic. These background signals usually have larger magnitude than signals from small cracks. Moreover, when the signal from a small crack is combined with the background signal, its shape is distorted by the slope of the background signal. This increases the difficulty of detecting small cracks.

To apply PCA to this problem, we noticed that there is a difference between the spatial variations of the small crack signal and the background signal. Signals from small cracks tend to have faster spatial variations. Also, because the crack signal has a rather small support, its energy is only a small portion of the overall energy of the image. Therefore, we can consider the crack signals as details of the image which can be separated from the background by using the PCA method.

Implementation of the PCA method

Based on the idea discussed above, the implementation of the method includes the following steps:

1. Divide the image into equally sized blocks. Usually a square block shape is used. Rearrange the data within each block to form a one-dimensional vector.
2. Use the vectors obtained from all the blocks as training data to train the Hebbian learning network. Before the training, the learning rate and the number of desired principal eigenvectors need to be decided. Either Oja's learning rule or the generalized learning rule can be used to update the weights. The network is said to have converged if the weights have very small changes during one epoch.
3. Use the trained network to obtain the principal eigenvectors and compute the expansion coefficient for each block and each eigenvector. Then the background signal in each block is reconstructed by using the eigenvectors and the expansion coefficients [51]. The one-dimensional reconstruction results are rearranged into two-dimensional blocks.
4. The reconstruction established up this step has a strong block effect due to the discontinuities at block boundaries. To remove the block effect, a moving-average smoothing is used. To save processing time, two one-dimensional smoothing operations (one in the horizontal direction, and the other in vertical direction) are used instead of one two-dimensional smoothing operation. The smoothing operations use rectangular smoothing windows and the window size is selected to be equal to the block size.
5. The background is then subtracted from the original data and another moving-average smoothing operation is applied to the subtraction result to remove spike-like noise. This noise usually comes from electronic interference and can be relatively large when compared to signals from small cracks.

Some further considerations

The major processing procedure of the PCA method is given in the above section. However, to generate an optimal result, some details of the method need to be considered. The first thing to be considered is how to select the processing parameters. The most important parameters are the block size and the number of principal eigenvectors used in reconstruction. The block size is important because it determines the spatial resolution of the reconstructed image. Therefore, we need to select a block size within which the background signal can be considered as a slow variation and the crack signals can be considered as fast variations. Because a crack signal generally has a length greater than its width, the block size can be selected to be close to the width of the crack signal. This is feasible because the width of the crack signal is basically determined by probe characteristics and remains almost a constant. The selection of the number of principal components is not very critical because experimental results indicate that the background obtained is not very sensitive to it. To save learning time, a smaller number of eigenvectors is desired. However, if this number is too small a significant part of the background signal will be lost in the reconstruction. A good selection can be made by computing a relatively large number of eigenvalues and then selecting those which are significantly larger than the rest of the eigenvalues.

Another problem that needs to be considered is that when there is a large crack signal in the data, the background signal may be distorted by the presence of a large flaw signal. This condition is shown by large expansion coefficients near the flaw position. To solve this problem, we used a statistical test to locate such points. First, for the expansion coefficients of each eigenvector, the difference between each pixel and the average of its eight neighbor

pixels is calculated. Then the mean and the standard deviation of this difference are calculated. Every expansion coefficient is then reexamined and those which differ from the mean by more than 3 times the standard deviation are considered as singular points. These singular points are considered to arise not as a result of the background signal, but as a result of some large crack signal. The value of the expansion coefficient at each singular point is then replaced by the average value of its eight neighbors.

Processing examples

The PCA method discussed above was tested by using inspection data obtained from an industrial standard MRPC probe. Due to space limit, we only give an example of the PCA background removal processing here. More processing results can be found in Reference [63]. The test sample in this example contains nine OD EDM slots in a tube with circumferential extent from 10 degrees to 90 degrees. The depth of the flaws is 20 percent of the tube wall thickness. The data was preprocessed using masked polynomial background removal and automatic phase adjustment. Fig. 8.1 shows the preprocessed data containing signals from the 10°, 20°, and 30° slots. The data was then processed by using the PCA method. The block size was selected as 16 by 16 and only three principal eigenvectors were used. The generalized learning rule was used to train the Hebbian learning neural network. After 20 epochs, the weight vectors of the network converged to the three principal eigenvectors of the covariance matrix of the input data. The three principal eigenvalues were 3120155, 932939, 406210, respectively. Some statistics of the three principal eigenvectors are shown in table 8.1.

Fig. 8.1. MRPC data of the three artificial flaws after preprocessing.

Table 8.1. Statistics of three principal eigenvectors.

	Mean	Minimum	Maximum
Eigenvector 1	0.061165	0.010311	0.081576
Eigenvector 2	0.014011	-0.081951	0.108136
Eigenvector 3	0.007834	0.007834	0.088679

The images of the three principal eigenvectors are shown in Figs. 8.2, 8.3, and 8.4. It is clear that eigenvector 1 contains most of the DC energy of the image. From the shapes of the eigenvectors, we can also see that eigenvectors 1 and 2 reflect the diagonal variations of the background signal in the input image. The eigenvalues of the three principal eigenvectors are much larger than the rest of the eigenvalues. This means that the three principal eigenvectors represent most of the energy in the input image.

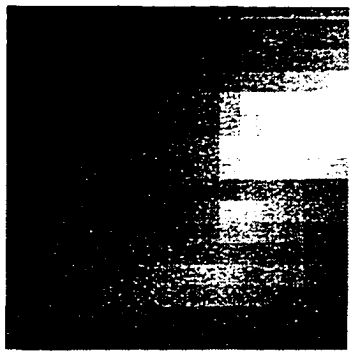


Fig. 8.2. Image of eigenvector 1 obtained using generalized Hebbian learning.

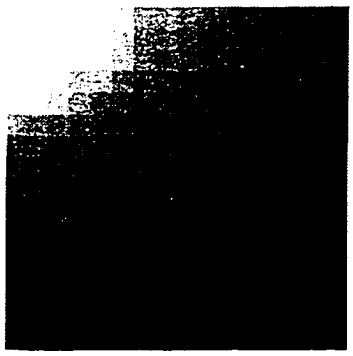


Fig. 8.3. Image of eigenvector 2 obtained using generalized Hebbian learning.

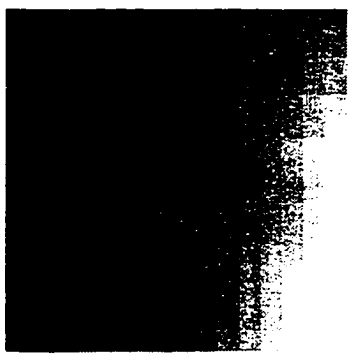


Fig. 8.4. Image of eigenvector 3 obtained using generalized Hebbian learning.

After the three principal eigenvectors were obtained, the background image was generated by reconstructing the input image using only the three principal components. The processing result was then obtained by subtracting a smoothed version of the background image from the input image. Fig. 8.5 illustrates the processing result of the PCA background removal method.

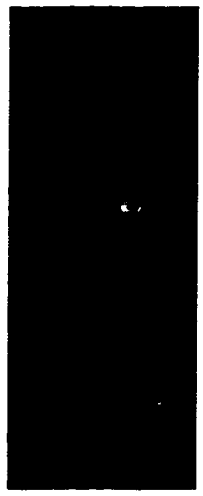


Fig. 8.5. Result image of the PCA background removal processing.

To illustrate the effect of the PCA method on small crack signals, we plotted the one-dimensional profiles of the signals for the 10°, 20° and 30° EDM slots before and after the processing. The three one-dimensional comparisons are shown in Figs. 8.6, 8.7, and 8.8, respectively. From these three comparisons we can see that the background signal has been significantly reduced, and the residue is constrained in a small dynamic range and is smaller than the signal from the 10° EDM slot. At the same time, the signal intensity remains almost unchanged for the 10° and 20° slots. However, there is a visible decrease in the signal

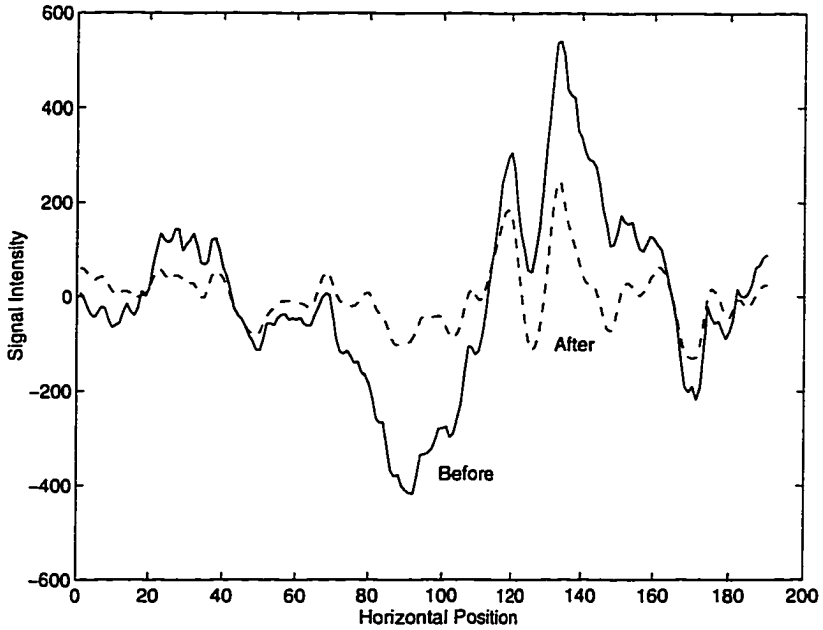


Fig. 8.6. Comparison of the horizontal profile of the signal for the 10° flaw before and after processing.

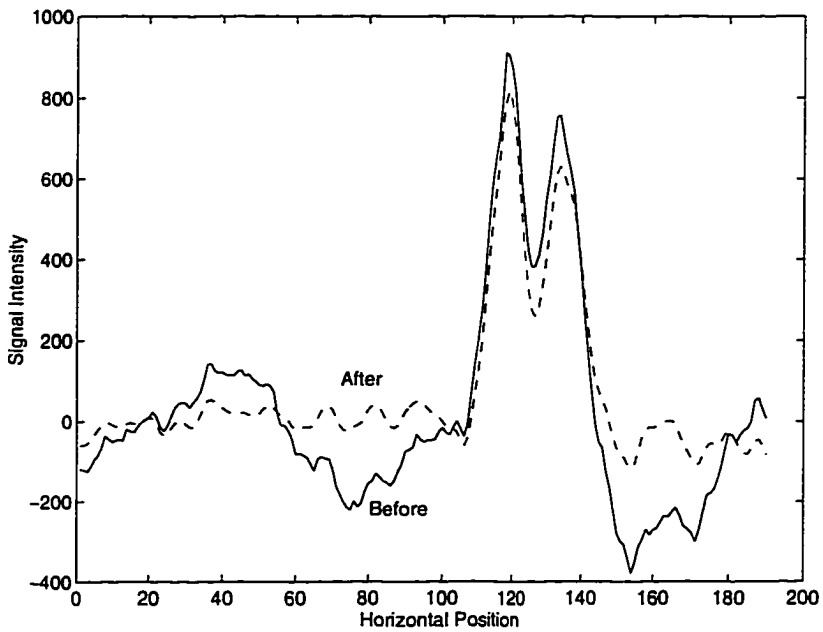


Fig. 8.7. Comparison of the horizontal profile of the signal for the 20° flaw before and after processing.

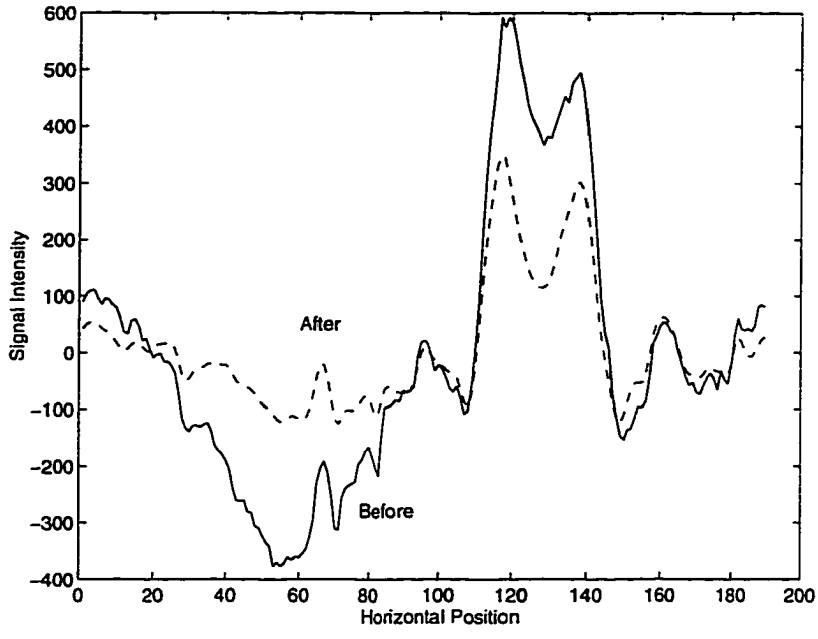


Fig. 8.8. Comparison of the horizontal profile of the signal for the 30° flaw before and after processing.

intensity of the 30° slot. The reason for this is that the two peaks of the eddy current signal of the 30° slot are starting to merge, which increases the spatial support of the flaw signal. In other words, it is more like the background signal than the signals from the 10° and 20° slots. Therefore, the increase of the signal-to-noise ratio for the 30° slot is not as large as that of the 10° and 20° slots.

Since two-dimensional eddy current data can be more clearly visualized in a wireframe display, we plotted the three-dimensional wireframe display for the signal from the 10° slot before and after processing as a comparison. Figs. 8.9 and 8.10 show the signal for the 10° slot before processing and after processing, respectively. As can be seen from these two plots, after processing the flaw signal is more symmetric because most of the

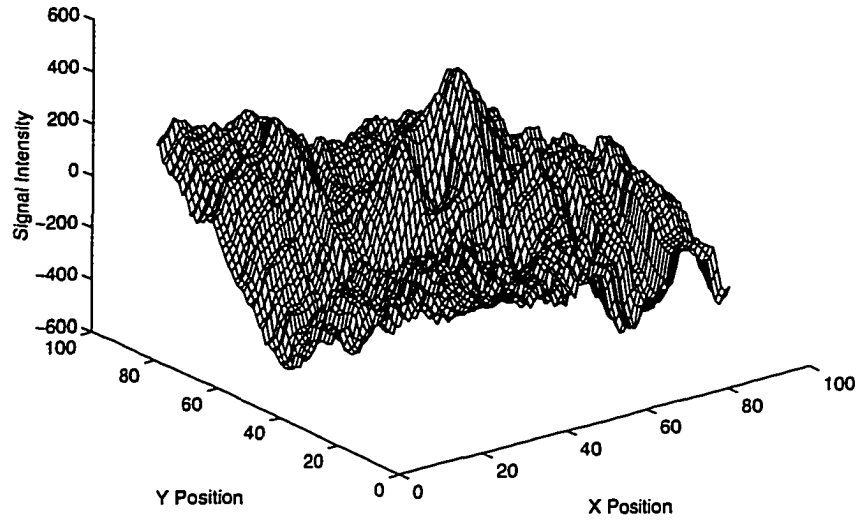


Fig. 8.9. Wireframe display of the signal for the 10° slot before processing.

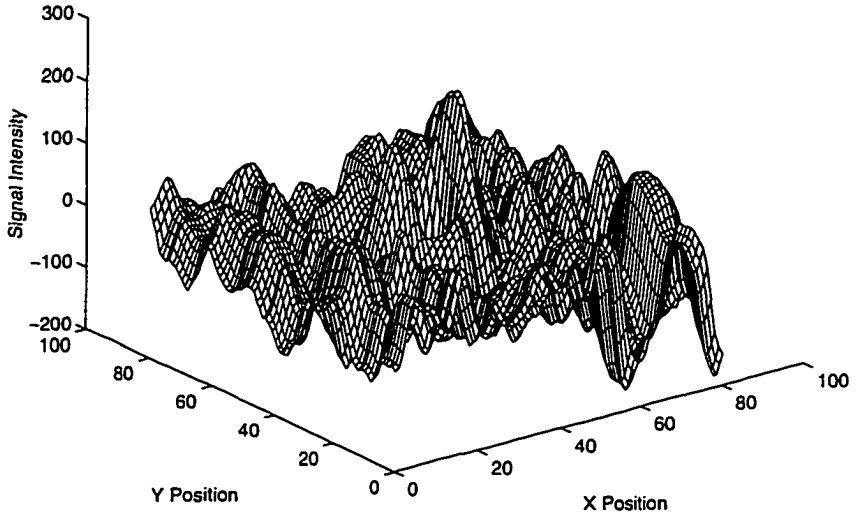


Fig. 8.10. Wireframe display of the signal for the 10° slot after processing.

background signal has been removed. It is clear that the PCA processing has good potential to increase the detectability of small cracks.

Conclusions

In this chapter we presented the principal component analysis method as an image processing technique to remove background signals in eddy current inspection data and to enhance signals from small cracks. Processing results from MRPC data have shown the high potential of the method. The background removal by PCA can achieve large improvement in SNR with very small distortion in the shape of the flaw signals with small support. Since the PCA method does not distort flaw signals, it can be used as a flaw enhancement method for human analysts. It can also be used as a preprocessing step for flaw detection methods. The disadvantage of the PCA method is that it is relatively inefficient for signals from large flaws which have relatively large spatial supports. However, this should not be a serious problem because signals from large flaws are generally stronger than background signals. Those signals can be readily detected even without the PCA processing.

CHAPTER 9. CIRCUMFERENTIAL CRACK DETECTION USING GRAYSCALE MATHEMATICAL MORPHOLOGY

Introduction

In the preceding chapters we have discussed preprocessing techniques and a background removal method based on principal component analysis for improving the signal quality of two-dimensional eddy current inspection data. After the inspection data is processed by these techniques, the flaw signals are enhanced and the probability of finding a certain flaw is increased. Even though the data quality is improved by these processing procedures so that human analysts can find flaws more easily, in industrial practice we would still like to have an automatic crack detection method which can locate possible flaw indications in the inspection data without manual inspection by human analysts. As mentioned before, in heat exchanger tubes circumferential cracks are of special interest. Therefore, in our study we focused on the automatic detection of circumferential cracks. In this chapter we discuss a detection method for circumferential cracks based on grayscale mathematical morphological operations. This detection method is based on two important shape characteristics of the signal from a circumferential crack obtained using a MRPC: the maximum vertical extent of the signal and the minimum horizontal extent of the signal. In this detection method, grayscale morphological operations are used to remove features that do not have these two characteristics while keeping the crack signals almost intact. Test results have shown this method can be used to effectively locate possible crack signals.

Grayscale Mathematical Morphological Operations

Mathematical morphology is an important area in nonlinear image processing [65, 69-76]. The fundamental philosophy of mathematical morphology is well explained in the following statement: what you see depends on what you want to look at. All the morphological operations involve the concept of a structuring element which relates the input and the output of the morphological operation. It is the structuring element that defines what you want to look at. Detailed discussions on grayscale morphology can be found in References [65], [69], and [70]. Here we give just a simple description on basic grayscale morphological operations.

There are four basic grayscale mathematical morphological operations: erosion, dilation, opening, and closing. The latter two operations are built upon the first two operations. As the name implies, grayscale erosion erodes the input image and generates an output with smaller support and peak values. The extent of the erosion and its exact impact on the input image are decided by the shape of the structuring element used in the erosion. Dilation is a dual operation of erosion that extends the support and peak values of the input image. Opening is a combination of erosion and dilation. It is an erosion followed by a dilation with the same structuring element. Opening is usually used to remove small bright point-like features. Closing is the dual operation of opening, and it is a dilation followed by an erosion with the same structuring element. Closing is usually used to fill small holes in the input image.

An important extension to standard mathematical morphological filtering is the so-called soft morphological filtering, or rank-order filtering. In standard morphological

filtering, the concept of set comparison is crisp. For example, if we say a structuring element is below the input image, it means that all of the pixels in the structuring element have a value smaller than the value of the corresponding pixel in the input image. In the rank-order filtering, the concept of set comparison is fuzzy. That is, we say that the input image matches the structuring element if at least some number of pixels, or some percentage of pixels, in the structuring element match that of the input image. Rank-order filtering is useful when there is noise in the data and the shape of the object to be detected is somewhat incomplete.

The Circumferential Crack Detection Method

Characteristics of the signal from a circumferential crack

A circumferential crack is considered a horizontal OD crack with a very small width (tight crack). The shape of the signal from a circumferential crack changes with the length of the crack. However, due to the tight opening of circumferential cracks and the axial symmetric geometry of the MRPC coil, signals from circumferential cracks share the following two important characteristics:

Maximum vertical extent

A circumferential crack signal has a limited vertical support. The signal is practically zero outside this vertical support. The actual length of this vertical limit is decided by the size of the coil and the skin depth at the test frequency. Usually, the vertical limit is only slightly larger than the outside diameter of the coil.

Minimum horizontal extent

A circumferential crack signal has a minimum horizontal extent. No signal from a circumferential crack can have a horizontal support smaller than a horizontal limit. The horizontal limit is decided by the size of the coil, the skin depth at the test frequency, and the minimum length of a physical circumferential crack. The horizontal minimum size is usually slightly larger than the vertical maximum size.

Procedure of the detection method

Based on the two characteristics of signals from circumferential cracks, we have developed a detection method based on grayscale morphological operations. The detection method is described as follows

1. Remove the null offset in the inspection data by using masked polynomial fitting or median filtering.
2. Calibrate the phase of the preprocessed data by using a standard calibration procedure or the automatic phase adjustment. Use only the vertical channel signal in later processing.
3. Remove background signals in the vertical channel data using the PCA method discussed in Chapter 8.
4. Smooth the data using a rectangular moving average window.
5. Open the data using a small flat structuring element. This step removes point-like noise in the data.
6. Close the data using a horizontal flat structuring element with a small length. This step is used to fill the valley between the two peaks for signals from short cracks.

7. Open the data using a horizontal flat structuring element with the a length defined by the minimum horizontal extent. This step eliminates any feature that has a horizontal support smaller than the minimum horizontal support.
8. Erode the data using a vertical flat structuring element with a length defined by the maximum vertical extent. This gives an image with all the signals from circumferential cracks removed. Signals having a vertical support larger than the maximum vertical extent remain in the result image.
9. Subtract the result image of step 8 from the result image of step 7. This gives an image which does not contain any signal with a small horizontal support or a large vertical support. In other words, all the signals remaining in this image must satisfy both the minimum horizontal extent requirement and the maximum vertical extent requirement.
10. Open the data using a small flat structuring element. This step is used to remove point-like noise that may result from the subtraction in step 9.
11. Threshold the data to create a binary crack map. The threshold must be selected properly to ensure high probability of detection without many false alarms. The threshold can be decided by calibrating the probe using artificial flaws with various sizes and depths. A critical flaw size and depth should be set up to derive the threshold.

Test Results

To illustrate the procedure of the crack detection method and to demonstrate the result of processing, we selected the inspection data of a tube containing nine circumferential EDM slots. The depth of the slots is 20% TW, and the circumferential extents vary from 10

degrees to 90 degrees. After truncating entry and exit points and interpolating and aligning in each row, the size of the raw data is 193 by 1351 pixels. The first step of processing is background removal using second-order masked polynomial fitting. The background removal was first applied to the horizontal direction, and then to the vertical direction. The margins were 20% for the horizontal direction and 5% for the vertical direction. The data was then processed by automatic phase adjustment. Fig. 9.1 shows the vertical channel signal after the automatic phase adjustment. The next step of the processing is to remove background signals with smaller spatial periods using the PCA method. The result of PCA processing is shown in Fig. 9.2. From Fig. 9.2 we can see that most background signals have been successfully removed. The result of Fig. 9.2 was smoothed by using an 8 by 8 moving

Fig. 9.1. Image of the vertical channel signal after preprocessing.

window, and the result is shown in Fig. 9.3. Then the data was opened with a 6 by 6 flat structuring element to remove point-like noise. Following this opening a closing with a 16 by 1 flat structuring element was performed. This step was used to fill the valley between the two peaks in the signals of a short crack. A second opening with a 25 by 1 flat structuring element was used to remove all features smaller than the minimum horizontal extent. The resulting image is shown in Fig. 9.4. To enforce the maximum vertical extent requirement, the image in Fig. 9.4 was eroded by using a 1 by 16 flat structuring element and the resulting image was subtracted from the image in Fig. 9.4. The sizes of the structuring elements in all the morphological operations were derived from the sizes of the coil and the test parameters.

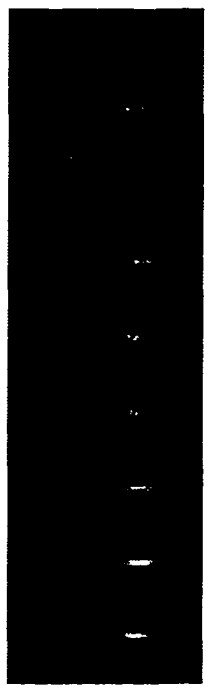


Fig. 9.2. Result image after PCA background removal.

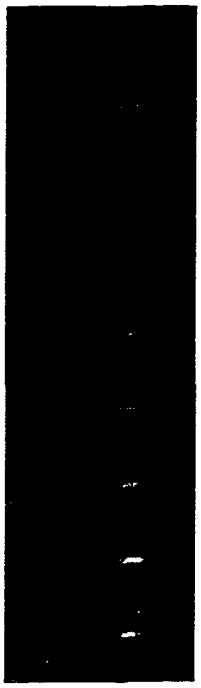


Fig. 9.3. Result image after smoothing by moving average.

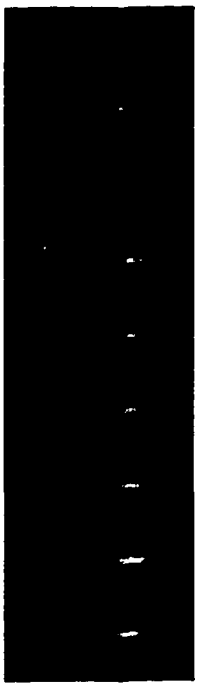


Fig. 9.4. Result image after the second opening.

Fig. 9.5 shows the result after using the subtraction and a further opening with a 6 by 6 flat structuring element to remove point-like noise. It can be seen from Fig. 9.5 that most of the noise signals have been removed. However, there are some small artifacts due to repetitive morphological operations. To generate the crack map, the grayscale image in Fig. 9.5 was then thresholded to create a binary image. The crack map is shown in Fig. 9.6. All nine EDM slots in the inspection data are shown clearly in the crack map, including the 10° slots which are hardly distinguishable in the original data. Due to the nonlinear nature of morphological operations and the diffusive nature of eddy current, the cracks shown in the crack map do not have the same shape as the physical flaws. However, the crack map still provides accurate positions of the cracks. The capability of locating potential flaws can be very helpful to human analysts when the quality of data is not very satisfying.

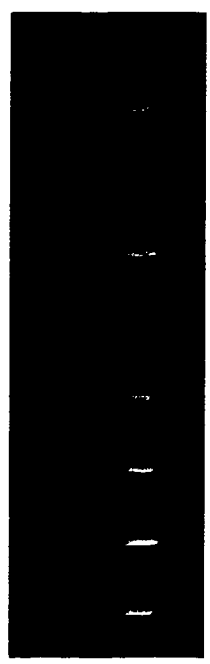


Fig. 9.5. Result image after the subtraction.

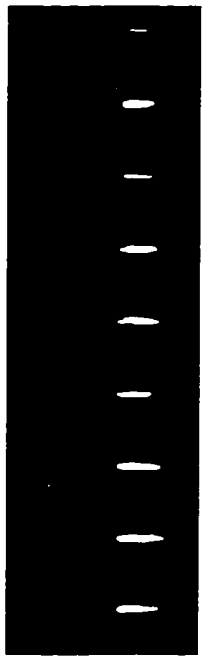


Fig. 9.6. The binary crack map.

Conclusions

In this chapter we discussed a detection method for circumferential cracks based on grayscale mathematical morphological operations. This detection method is based on two important shape characteristics of the signal from a circumferential crack obtained using a MRPC: the maximum vertical extent of the signal and the minimum horizontal extent of the signal. Test results have shown this method can be effectively used to locate possible crack signals in low SNR inspection data. This method can provide significant assistance to human analysts in locating possible flaw indications. However, due to the distortion introduced by morphological operations, this method does not generate accurate flaw shape for the purpose of flaw characterization. In the following chapters we discuss image restoration methods which can be used to find the actual flaw shape from blurred impedance change images.

CHAPTER 10. LINEAR EDDY CURRENT IMAGE RESTORATION USING WIENER FILTERING

Introduction

In this chapter we discuss an image restoration method for two-dimensional eddy current data using the Wiener filtering technique. This method was originated from the understanding we obtained in studying the eddy current inverse problem for a small spherical flaw [77]. We based our inverse method on an approximate linear eddy current forward model. This forward model is based on the Born approximation and the reciprocity impedance change formula (equation (2.1)). From this linear forward model, the point spread function of the MRPC coil can be computed and the Wiener filtering method can be applied to restore the flaw image. We developed two approximation methods to compute the point spread function for the Wiener filtering method. Testing of the Wiener filtering restoration method included both synthesized and experimental data. We also studied the effect of noise on the restoration result and the limits of the linear approximation. Test results have shown it is possible to improve flaw characterization capability under certain conditions.

Formulation of the Linear Deconvolution Model

The linear forward model

We used an approximate linear forward model for the Wiener filtering eddy current image restoration. Here we briefly describe the formulation of the linear forward model

based on the Born approximation. The impedance change of an air-core coil placed on a testpiece can be computed by using the following reciprocity formula (equation (2.1))

$$\Delta Z = -\frac{1}{I^2}\int_{V'} \delta\sigma(\mathbf{r}')\mathbf{E}^0(\mathbf{r}')\cdot\mathbf{E}(\mathbf{r}')dv', \tag{10.1}$$

where I is the current in the coil, $\delta\sigma(\mathbf{r}')$ is the conductivity change in a material caused by a defect, $\mathbf{E}^0(\mathbf{r})$ is the incident electric field, and $\mathbf{E}(\mathbf{r})$ is the total incident field. The incident field is a function of the test coil, test frequency, and test geometry; but it is not a function of flaw conductivity distribution. The total field is a more complex function dependent on both the incident field, test parameters and the flaw conductivity distribution. The relationship between the incident field, the total field, and the flaw conductivity distribution can be expressed in the following volume integral equation (equation (3.10))

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) + \int_{V'} \delta\sigma(\mathbf{r}')\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{E}(\mathbf{r}')dv', \tag{10.2}$$

where $\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')$ is the Green's function for the test geometry. Since the total field is a function of both the incident field and the conductivity change, the impedance change ΔZ is a nonlinear function of the conductivity change $\delta\sigma(\mathbf{r}')$. To linearize the forward model for the purpose of image restoration, we can use the Born approximation which lets the total field be equal to the incident field

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}). \tag{10.3}$$

Now the impedance change becomes

$$\Delta Z = \frac{-1}{I^2}\int_{V} \delta\sigma(\mathbf{r})\mathbf{E}^0(\mathbf{r})\cdot\mathbf{E}^0(\mathbf{r})dv, \tag{10.4}$$

which is linear in the conductivity change. To build a two-dimensional image restoration model, we assumed that the conductivity change is uniform along the z direction, e.g.,

$$\delta\sigma(\mathbf{r}') = \delta\sigma(x', y'). \tag{10.5}$$

Now the impedance change can be expressed as

$$\Delta Z = K\int_{S'} \delta\sigma(x', y')\int_{z_1}^{z_2} \mathbf{E}^0(\mathbf{r}')\cdot\mathbf{E}^0(\mathbf{r}')dz'ds', \tag{10.6}$$

where $ds' = dx'dy'$ and we use K to denote a constant before the integration which is not of much importance here. Letting

$$g(x', y') = \int_{z_1}^{z_2} \mathbf{E}^0(\mathbf{r}')\cdot\mathbf{E}^0(\mathbf{r}')dz', \tag{10.7}$$

we can express the impedance change at a given point as

$$\begin{aligned}\Delta Z(x, y) &= K\iint \delta\sigma(x', y')g(x'-x, y'-y)dx'dy' \\ &= K\iint \delta\sigma(x', y')g'(x-x', y-y')dx'dy',\end{aligned} \tag{10.8}$$

where $g'(x, y) = g(-x,-y)$. We see that the impedance change image is the result of a linear convolution of the conductivity change image and the point spread function (PSF) $g'(x, y)$. The point spread function can be computed if the incident field distribution is known, and the incident field can be obtained either by using Dodd and Deeds' formula [28] or using photoinductive measurements. We also developed approximation methods to compute the point spread function for tube geometry.

The Wiener filtering method

Given the convolution formula above, it is straightforward to apply an inverse filtering to restore the flaw image from the impedance image since we all know that the

convolution in spatial domain is multiplication in Fourier domain. However, the simple inverse filtering approach does not work very well when the Fourier transform of the point spread function has some small values, and when there is additive noise in the input. In these cases, Wiener filtering [78] is a better approach which can restore images in the presence of blur as well as noise. With the presence of noise, the linear observation model can be rewritten as

$$i(x,y) = f(x,y) * g'(x,y) + n(x,y), \tag{10.9}$$

where $i(x,y)$ is the impedance change image obtained from measurements, $f(x,y)$ is the flaw image, $g'(x,y)$ is the point spread function, and $n(x,y)$ is the additive noise.

To reduce the effect of noise on inversion, the Wiener filtering minimizes the mean square error between the restoration and the ideal result by

$$\varepsilon^2 = E\left\{\left|f(x,y) - \hat{f}(x,y)\right|^2\right\}, \tag{10.10}$$

where $\hat{f}(x,y)$ is an estimation of the flaw image. For a linear deconvolution model, we assume that this estimation is the result of the linear convolution of the impedance measurement and the Wiener filter by

$$\hat{f}(x,y) = h(x,y) * i(x,y), \tag{10.11}$$

where $h(x,y)$ is the impulse response of the Wiener filter. This linear mean square problem satisfies the well known orthogonality condition of

$$E\left\{\left[f(x,y) - \hat{f}(x,y)\right]^* i(x,y)\right\} = 0. \tag{10.12}$$

Substituting the definition for $\hat{f}(x,y)$ into the above equation, and assuming that the noise is independent of the flaw image, we have

$$R_{ii}(x,y) * h(x,y) = R_{if}(x,y). \tag{10.13}$$

Where $R_{ii}(x,y)$ is the autocorrelation function of $i(x,y)$, and $R_{if}(x,y)$ is the crosscorrelation function of $i(x,y)$ and $f(x,y)$. Transforming this to the Fourier domain gives

$$H(u,v) = \frac{S_{if}(u,v)}{S_{ii}(u,v)}, \tag{10.14}$$

where $S_{if}(u,v)$ and $S_{ii}(u,v)$ are the Fourier transforms of $R_{if}(x,y)$ and $R_{ii}(x,y)$, respectively. From the definition of $i(x,y)$, we know that

$$S_{if}(u,v) = G'^{*}(u,v)S_{ff}(u,v), \tag{10.15}$$

where $G'(u,v)$ is the Fourier transform of $g'(u,v)$, and

$$S_{ii}(u,v) = |G'(u,v)|^2 S_{ff}(u,v) + S_{nn}(u,v), \tag{10.16}$$

where $S_{ff}(u,v)$ is the power spectrum of the flaw image, and $S_{nn}(u,v)$ is the power spectrum of the additive noise. Thus from equation (10.14) the general form of the Wiener filter is

$$H(u,v) = \frac{G'^{*}(u,v)S_{ff}(u,v)}{|G'(u,v)|^2 S_{ff}(u,v) + S_{nn}(u,v)}. \tag{10.17}$$

From this equation, we see that to apply the Wiener filter, we need to know the power spectra of the signal and the noise. In real applications, this information generally is not available.

But if we can assume that the noise is somewhat white, then we can use the following simplified form of the Wiener filter,

$$H(u,v) = \frac{G'^*(u,v)}{|G'(u,v)|^2 + \alpha}, \tag{10.18}$$

where α is a parameter related to the signal-to-noise ratio of the input data. There is a tradeoff in the selection of α. For a large value of α, the noise is suppressed, but the reconstructed image will tend to be blurred. For a small value of α, the reconstructed image is sharp, but the noise will tend to be large.

The point spread function

It can be seen from the above discussion that the computation of the PSF requires knowledge of the incident field for the given coil and testpiece. For half space and thin plate geometry, the incident field distributions can be directly obtained using Dodd and Deeds' results [28]. However, in steam generator tubing inspections, the geometry is a pancake coil over the internal surface of the tube. The exact theoretical solution of the incident field for this geometry is not yet available. The difficulty with the analytical solution is mainly due to the lack of symmetry in this geometry. To deal with this problem, we developed two approximation methods to compute the point spread function for tube geometry. The two approximations are the flat plate approximation and the projection approximation.

The flat plate approximation

The flat plate approximation replaces the tube wall with a flat plate with the same thickness and uses the incident field for the flat plate to approximate the incident field in the tube wall. The approximation is illustrated in Fig. 10.1 (dotted lines). In this approximation,

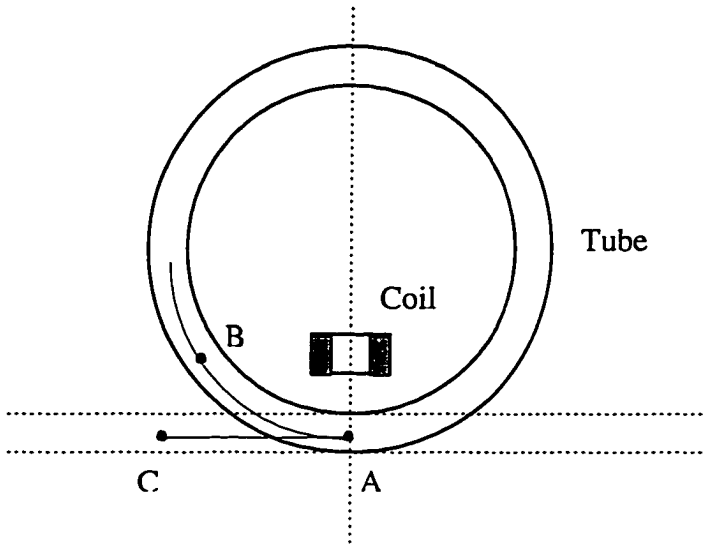


Fig. 10.1. Flat plate approximation to estimate PSF in tube geometry.

for a point B in tube wall, its incident field is estimated by using the incident field of a point C in the flat plate. Points B and C have the same depth under the tube wall internal surface and the flat plate top surface, respectively. Also, the arc length AB is equal to the distance AC.

The incident field of a cylindrical air core coil at a point in a thin plate is given by Dodd and Deeds [28] as

$$\mathbf{E}(\mathbf{r}) = T\int_{-\infty}^{\infty} \frac{I(\alpha)(e^{-\alpha l_1} - e^{-\alpha l_2})\left[(\alpha+\alpha')e^{-\alpha' z} + (\alpha'-\alpha)e^{-\alpha'(2H-z)}\right]}{\alpha\left[(\alpha+\alpha')^2 - (\alpha'-\alpha)e^{-2\alpha' H}\right]} d\alpha. \qquad (10.19)$$

Detailed explanation of this equation and the numerical method for its calculation can be found in Chapter 4.

Projection approximation

From Fig. 10.1, it can be easily seen that the thin plate approximation is valid only if the coil is much smaller than the tube; thus, the tube surface near the coil can be considered as sufficiently flat. When the diameter of the coil is comparable to the inner radius of the tube, the flat plate approximation fails to give an accurate PSF. To deal with this problem, we considered another approximation - the projection approximation as shown in Fig. 10.2.

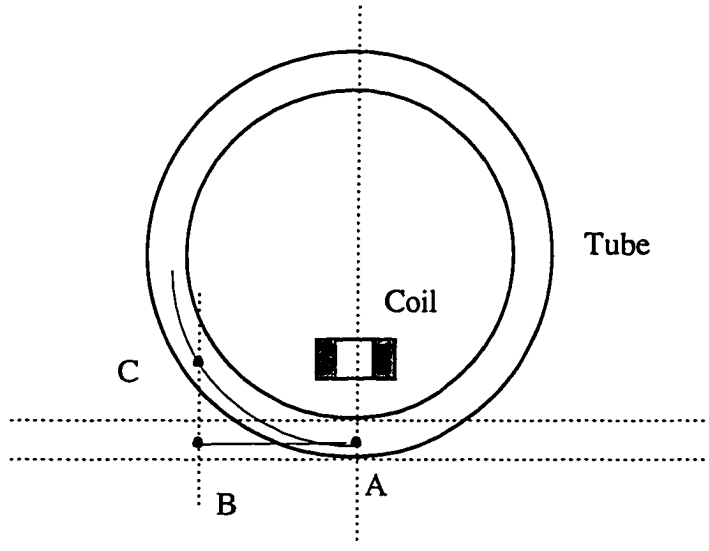


Fig. 10.2. Projection approximation to estimate the PSF in tube geometry.

In the projection approximation, the incident field of a point B in the tube wall is approximated by the incident field of a point C in the thin plate. The distance between B and tube internal surface is the same as the distance between C and the flat plate top surface. The difference of this approximation with the flat plate approximation is that point C is now on

the projection line of point B; thus, B and C have equal horizontal distances to the center line of the coil.

The projection approximation is based on the observation that when the curvature of the tube internal surface is large, the field intensity of a point in the tube wall is not proportional to the arc length between the point and the coil center line; but, more likely it is proportional to the horizontal distance between the point and the coil center line. However, our experience has shown that for most industrial standard tubes and coils, the flat plate approximation and the projection approximation give very similar results. Therefore, the projection approximation is necessary only when a relatively large coil or a relatively small tube is used.

Test Results from Synthesized Data

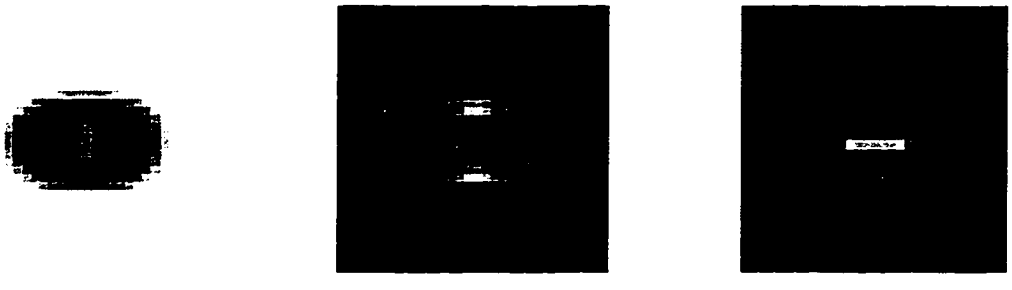
The effect of noise

To investigate the feasibility of the Wiener filtering method and to find some qualitative properties of the eddy-current inverse problem, we first applied the method to synthesized impedance change data given by the volume integral method code. The flaw used was an 8 mm long, 0.5 mm wide and 0.5 mm deep slot on the bottom side of a 1 mm thick aluminum plate. When the radius of the probe is much smaller than the tube inner radius, the thin plate setting is a good approximation for testing a circumferential OD flaw with a MRPC probe. The probe used was a 504-turn air-core coil with an inner radius of 3.8 mm and an outer radius of 5.635 mm. The height of the coil was 2.4 mm and the liftoff was 0.229 mm. The first test used a frequency of 10 kHz which corresponds to a skin depth of

(a) Real part (b) Imaginary part

Fig. 10.3. Point spread function for the synthesized data.

1.17 mm. The real part image and the imaginary part image of the point spread function are given in Fig. 10.3(a) and Fig. 10.3(b), respectively. The point field function was computed using equation (10.7) and the incident field was computed using equation (3.33). All the images shown have 64 by 64 pixels with a grid separation of 0.5 mm. The real part image and the imaginary part image of the synthesized impedance change obtained using the volume integral method with wavelet expansion are shown in Fig. 10.4(a) and Fig.10.4(b). Using equation (10.18), the flaw image was restored and is shown in Fig. 10.4(c). Due to an



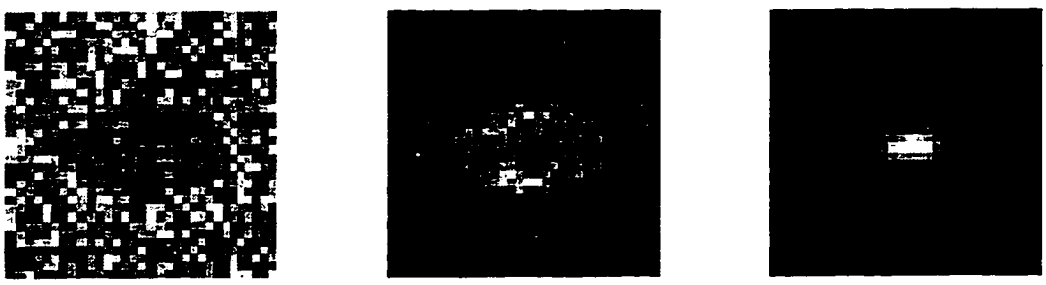
(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 10.4. Test results for the no noise case.

error introduced by the linearization, the restored flaw image is not uniform in intensity. However, it still gives a rather sharp image of the original flaw. To test the effect of the noise in the inverse process, white Gaussian noise was added to the impedance change image. A test image with 20 dB SNR and a test image with 10 dB SNR were used. The SNR used here is defined as the ratio of the total power of the signal and the total power of the noise in the whole image. The real part images and imaginary part images of these two data sets are shown in Fig. 10.5(a), Fig. 10.6(a), Fig. 10.5(b), and Fig. 10.6(b), respectively. The result images from the Wiener filtering are shown in Fig. 10.5(c) and Fig. 10.6(c), respectively.

(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 10.5. Test results for the 20 dB case.



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 10.6. Test results for the 10 dB case.

Notice that here α has been adjusted according to the signal-to-noise ratio to obtain a good result. For both cases in the result images the flaw region is still visible. However, due to the presence of noise, the reconstructed image has more noise and there is more distortion in the flaw shape as SNR decreases.

Result for compactly spaced flaws

To verify the applicability of this method in the case of multiple flaws, we tested it with different flaw settings. In this case, two horizontal cracks were placed one next to the other in the vertical direction. Each flaw had a length of 8 mm, width of 0.5 mm, and depth of 0.5 mm. Other test settings were the same as before. The frequency was 10 kHz, which corresponds to a skin depth of 1.17 mm. The distance between the edges of the two cracks was 1.5 mm. The SNR was 10 dB. The real part impedance change image, the imaginary part impedance change image, and the result image are shown in Fig. 10.7. We can see that the two cracks are well separated in the restoration result. Our other tests [63] on multiple flaws restoration have revealed the fact that the resolution capability of eddy current testing is

(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 10.7. Test results for the two cracks case.

limited by the skin depth. Therefore, flaws with a distance several times smaller than the skin depth cannot be separated successfully.

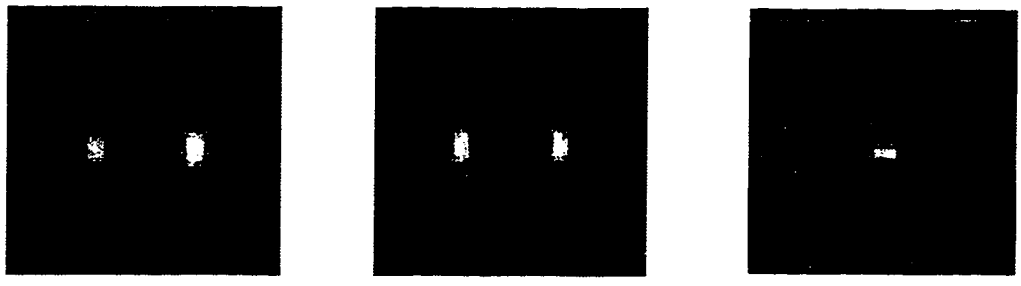
Limitations of the approximate linear deconvolution model

We performed other tests to evaluate the limitations of the approximate linear deconvolution model. Due to space limit, results of those tests are not included. Interested readers can find details on these tests in Reference [63]. Here we only summarize several observations from those tests:

1. The resolution of eddy current testing is proportional to the skin depth of the test.
2. The intensity of the restored flaw is not uniform due to the approximation error of the linear forward model.
3. When the flaw is large as compared to the coil or when the flaw has high aspect ratio, the linear approximation model fails to give good results due to its incapability of modeling the strong nonlinear interaction between the flaw and the coil in those cases.

Test Result for Experimental Data

To verify the validity of using this method in real applications, we tested it with experimental data. In this test, a 2-D scan was taken on a sample of a 4.204 mm long, 0.210 mm wide, and 0.916 mm deep crack on the surface of a thick Al6061 plate. The plate was thick enough to be considered as a half space. The probe used was an air-core coil with an inner radius of 3.8 mm, outer radius of 5.635 mm, and a height of 2.4 mm. The liftoff was 0.379 mm. The test frequency was chosen as 50 kHz which corresponds to a skin depth of



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 10.8. Test results for the experimental data.

0.449 mm. The 2-D scan was taken on a grid of 64 by 64 data points, with grid separation of 0.4 mm in both horizontal and vertical directions. The crack was laid in the horizontal direction. Before the linear deconvolution processing, the experimental data were preprocessed to remove the background signal introduced by the change of liftoff during scanning. Since the surface of the aluminum plate is not perfectly flat, there is a rather large liftoff change along the vertical direction. The real part image and the imaginary part image of the preprocessed impedance change measurement and the reconstruction result are shown in Fig. 10.8. In the reconstructed image, a horizontal flaw is clearly visible. The horizontal extent of the reconstructed flaw is about 20% smaller than the exact length of the crack. The reconstructed flaw is somewhat blurred along the vertical direction. But overall there is no significant difference between the actual flaw shape and experimental result. This means that the linear deconvolution method can be used to process real-world inspection data as well, if the quality of the data is carefully controlled.

Conclusions

In this chapter we developed a two-dimensional linear approximate observation model for the interaction between the flaw image and impedance change image based on the Born approximation. With this linear imaging model, we applied the Wiener filter method to restore the flaw image from a noisy measurement of the impedance change image. The linear model and the restoration method were tested and evaluated by using both synthesized and experimental data. The effect of the noise and the skin depth was analyzed by using synthesized data. It was found that the Wiener filter method is relatively robust even when the SNR of the test data is as low as 10-20 dB. Based on the observation of several test cases, we found that the theoretic resolution limit of the eddy-current testing is basically determined by the skin depth. The method has also been tested for the case of compactly distributed flaws. The results indicate that it is capable of separating flaws which are not extremely close together in terms of skin depth. Testing of the method under an extreme condition revealed the limit where the linear approximation breaks down. The experimental results also proved the validity of the method under the assumption that the SNR of the measurement data is acceptable.

CHAPTER 11. LINEAR EDDY CURRENT IMAGE RESTORATION USING MAXIMUM ENTROPY METHOD

Introduction

One important problem in MRPC data processing is restoring a flaw image from a measured impedance change image. In the previous chapter, we applied the Wiener filter method to the MRPC image restoration problem. In this chapter we present a new technique - the maximum entropy method (MEM) for eddy current image restoration. Part of the content discussed in this chapter has been presented in Reference [79]. MEM has been successfully applied to many problems, such as spectral analysis, image restoration, and image reconstruction from projections [80-84]. In the area of image restoration, MEM has found many applications in radio astronomy [82]. Many studies of MEM have shown its superior performance over other image restoration techniques in both resolution, dynamic range, and noise reduction. In this chapter we discuss how to apply MEM to the eddy current image restoration problem. We will also give some test results of MEM and compare the results of MEM and the Wiener filter method.

The Maximum Entropy Method

Based on the linear forward model developed in the last chapter, we can restore the flaw image by using simple inverse filtering which is basically a linear deconvolution. However, for MRPC eddy current inspection data, the direct deconvolution by inverse

filtering is not practical due to its sensitivity to noise. To model the effect of noise in the restoration process, we can use the following noisy linear observation model

$$i(x, y) = f(x, y) * g'(x, y) + n(x, y), \tag{11.1}$$

where $i(x, y)$ is the noisy impedance change measurement, $f(x, y)$ is the flaw image, $g'(x, y)$ is the point spread function, and $n(x, y)$ is the additive noise. The additive noise $n(x, y)$ is usually considered to be white and Gaussian. In the case of Wiener filtering, the objective of image restoration is to find an estimate of $f(x, y)$ so that the mean square error (MSE) between the true flaw image and the estimated flaw image is minimized. However, because of the ill-posedness of the eddy current inverse problem, the MSE criterion of Wiener filtering does not guarantee an optimal result. In MEM, the entropy of the restored flaw image is also used to regularize the ill-posed inverse problem. The entropy function of the restored flaw image can be expressed as [80]

$$S = -\sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \bar{f}(x, y) \log(\bar{f}(x, y) / \bar{m}(x, y)), \tag{11.2}$$

where

$$\bar{f}(x, y) = \hat{f}(x, y) / \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \hat{f}(x, y), \tag{11.3}$$

is the normalized positive input image, and

$$\bar{m}(x, y) = m(x, y) / \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} m(x, y), \tag{11.4}$$

is a normalized version of $m(x, y)$ which represents our prior information about the restoration. The principle of maximum entropy requires maximizing the entropy of the restored image. If we have no prior information about the flaw image, then from equation

(11.2) we can see that the image which maximizes the entropy is a constant image. Therefore, the MEM principle is actually a smoothing constraint which forces the restored image to contain a smaller noise component. Besides maximizing the entropy function, we also want to minimize the mean square error between the impedance change measurements and the predicted impedance change from an estimate $\hat{f}(x,y)$ of the unknown flaw image, which is given by

$$E = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \left| i(x,y) - (\hat{f} * g')(x,y) \right|^2 . \tag{11.5}$$

The objective of MEM is to minimize the cost function

$$Q = E - \alpha S, \tag{11.6}$$

where α is a regularization constant which controls the smoothness of the result. Since the entropy function is a nonlinear function, nonlinear optimization algorithms must be used to minimize the cost function. In this work we use the modified Newton-Raphson method [83], which iteratively updates the restoration based on the equation

$$\hat{f}^{(n)}(x,y) = \hat{f}^{(n-1)}(x,y) - \gamma_n \left(\frac{\partial Q / \partial \hat{f}(x,y)}{\partial^2 Q / \partial \hat{f}(x,y)^2} \right)^{(n-1)}, \tag{11.7}$$

where γ_n is the step size used to control the convergence speed. The restoration result is initialized to a constant image using a small positive constant. To apply the Newton-Raphson method, the first and second derivatives of Q must be known. The first and second derivatives of the mean square error can be shown to be

$$\frac{\partial E}{\partial \hat{f}(x,y)} = -2\,\mathrm{Re}\left\{ ((i - \hat{f} * g') * g^*)(x,y) \right\}, \tag{11.8}$$

and

$$\frac{\partial^2 E}{\partial \hat{f}(x,y)^2} = 2\sum_{l=0}^{N_x-1}\sum_{m=0}^{N_y-1}\left|g'(l,m)\right|^2 . \tag{11.9}$$

The first and second derivatives of the entropy function can be shown to be

$$\frac{\partial S}{\partial \hat{f}(x,y)} = -\left(\log\frac{\bar{f}(x,y)}{\overline{m}(x,y)} + \frac{1}{\ln 10}\right)\frac{\partial \bar{f}(x,y)}{\partial \hat{f}(x,y)}, \tag{11.10}$$

and

$$\frac{\partial^2 S}{\partial \hat{f}(x,y)^2} = -\frac{1}{\bar{f}(x,y)\ln 10}\left(\frac{\partial \bar{f}(x,y)}{\partial \hat{f}(x,y)}\right)^2 - \left(\log\frac{\bar{f}(x,y)}{\overline{m}(x,y)} + \frac{1}{\ln 10}\right)\frac{\partial^2 \bar{f}(x,y)}{\partial \hat{f}(x,y)^2}, \tag{11.11}$$

where

$$\frac{\partial \bar{f}(x,y)}{\partial \hat{f}(x,y)} = \frac{1}{\sum_x\sum_y \hat{f}(x,y)} - \frac{\hat{f}(x,y)}{\left(\sum_x\sum_y \hat{f}(x,y)\right)^2}, \tag{11.12}$$

and

$$\frac{\partial^2 \bar{f}(x,y)}{\partial \hat{f}(x,y)^2} = -\frac{2}{\left(\sum_x\sum_y \hat{f}(x,y)\right)^2} - \frac{2\hat{f}(x,y)}{\left(\sum_x\sum_y \hat{f}(x,y)\right)^3}. \tag{11.13}$$

Due to the maximum entropy method being an iterative method, it has a much higher computational complexity than the Wiener filter method. In each iteration, the most time consuming computation is to find the first derivative of the mean square error. The direct computation of convolutions in the spatial domain is very inefficient. When the size of the flaw image is a power of 2, we can speed up the calculation by using the FFT and the frequency domain equivalent of the convolutions in equation (11.8) to obtain

$$p_E(u,v) = [i(u,v) - \hat{f}(u,v)g'(u,v)]g'^{*}(u,v). \qquad (11.14)$$

Implementation of the MEM Image Restoration Method

For the eddy current image restoration problem, the implementation of the MEM approach can be summarized as:

1. Initialize the flaw image to a constant image by using a small positive integer for each pixel. This is the initialized $\hat{f}(x,y)$.
2. Compute the Fourier transforms of $i(x,y)$ and $g'(x,y)$ by using a 2-D FFT.
3. Compute the second derivative of the mean square error by using equation (11.9).
4. Compute the Fourier transform of $\hat{f}(x,y)$ by using a 2-D FFT.
5. Compute the Fourier transform of the first derivative of the mean square error by using equation (11.14) and then using a 2-D inverse FFT to obtain the first derivative of the mean square error.
6. Compute the first derivative of the entropy function by using equations (11.10), (11.3), and (11.4).
7. Compute the second derivative of the entropy function by using equations (11.11) and (11.13).
8. Select a proper γ_n and update the estimated flaw image by using equation (11.7), where the first derivative and the second derivative of the cost function are the sum of the first derivatives of the mean square error and the entropy function and the sum of the second

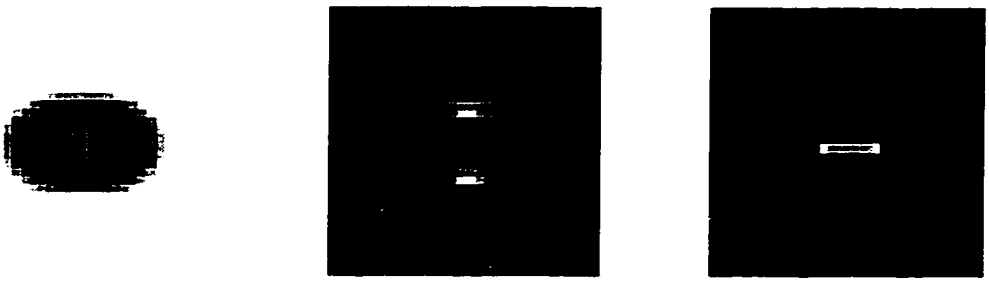
derivatives of the mean square error and the entropy function, respectively. The smoothing constant α should be selected according to the SNR of the data.

9. If the iteration number is larger than a predefined value or the change in cost function is smaller than a predefined value, stop and output the estimated flaw image. Otherwise, increment the iteration number and go to 4.

Test Results

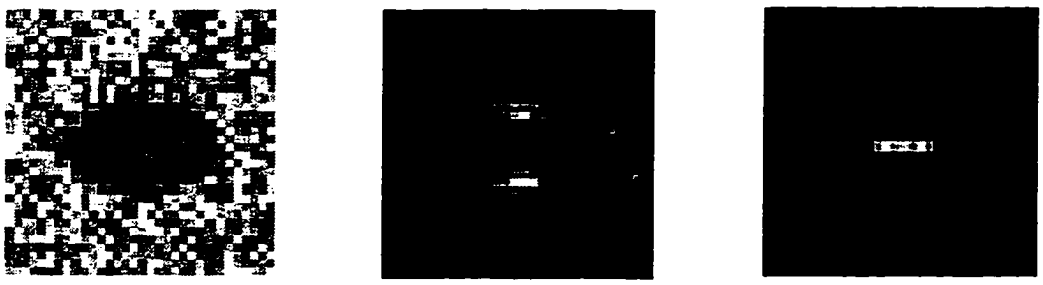
Test results for synthesized data

The maximum entropy method shown above has been tested using both synthesized data obtained from a volume integral forward model and experimental data from a surface-breaking crack on an aluminum slab. For all of the tests presented in this section, the step size γ_n was 0.001 and the smoothing constant α was 0.1. For synthesized data, a few cases were tested. In the first test, the effect of white Gaussian noise on the restoration result was studied. A rectangular flaw with a length of 8 mm, a width of 1 mm, and a depth of 1 mm was used in this test. Three cases were tested: no noise, SNR=20 dB, and SNR=10 dB. The real part image and the imaginary part image of the impedance change for the no noise case are shown in Fig. 11.1(a) and Fig. 11.1(b), respectively. The restoration result of the no noise case is shown in Fig. 11.1(c). The real part image and the imaginary part image of the impedance change for the 20 dB case are shown in Fig. 11.2(a) and Fig. 11.2(b), respectively. The restoration result of the 20 dB case is shown in Fig. 11.2(c). The real part image and the imaginary part image of the impedance change for the 10 dB case are shown in Fig. 11.3(a)



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 11.1. Test results for the no noise case.



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 11.2. Test results for the 20 dB case.

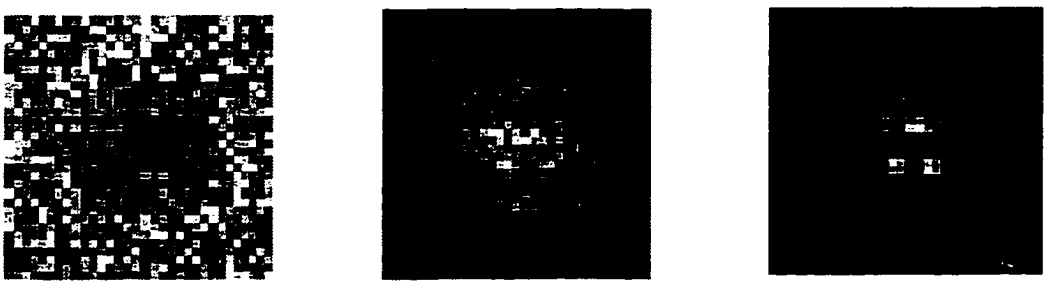
(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 11.3. Test results for the 10 dB case.

and Fig. 11.3(b), respectively. The restoration result of the 10 dB case is shown in Fig. 11.3(c). Comparing these results with the results from Wiener filtering (Fig. 10.4 - Fig. 10.6), it is clear that the MEM restoration has a better noise reduction capability and better resolution than the Wiener filtering restoration.

Test result for closely spaced flaws

The second test studied the capability of the restoration techniques to separate two closely spaced cracks. The flaw configuration for this test was two flaws with the same size as the one used in the first test, but with a separation of 4 mm between the two flaws. The SNR of the impedance change data is 10 dB. The real part image and the imaginary part image of the impedance change are shown in Fig. 11.4(a) and Fig. 11.4(b), respectively. The restoration result from MEM is shown in Fig. 11.4(c). Comparing the result from MEM with the result from Wiener filtering (Fig. 10.7), MEM again shows better performance over the Wiener filtering method. There is much smaller noise and a clearer gap between the two restored flaws.

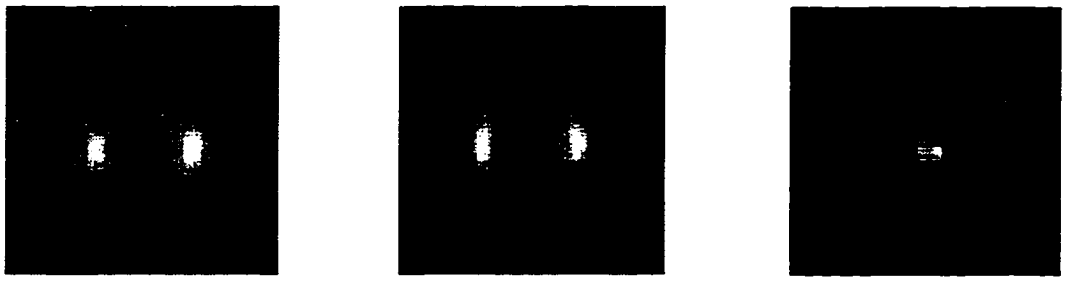


(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 11.4. Test results for the two cracks case.

Test Result for Experimental Data

The experimental data is the 2-D impedance change of an air-core coil over a surface-breaking crack with a length of 4.204 mm, a width of 0.210 mm, and a depth of 0.916 mm. The real part image and the imaginary part image of the impedance change measurement after background removal and smoothing are shown in Fig. 11.5(a) and Fig. 11.5(b), respectively. The restoration result from MEM is shown in Fig. 11.5(c). The restoration result from the Wiener filtering method (Fig. 10.8) is more blurred due to the strong background noise in the result image. It can be seen that the restoration result of MEM is much better, although the width of the crack is expanded due to experimental error and the error introduced by the linear approximation.



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 11.5. Test result for the experimental data.

Conclusions

In this chapter we applied the maximum entropy method to the eddy current image restoration problem. The MEM image restoration technique developed here is based on an approximate linear forward model obtained using the reciprocity impedance change formula

and the Born approximation. The cost function of MEM comprises of two parts: the first part is, the mean square error which requires the estimated flaw image to be close to the true flaw image; and, the second part, is the entropy function which requires the estimated flaw image to be smooth. We used a modified Newton-Raphson method to minimize the cost function and derived the first and second derivatives of the mean square error and the entropy function, which are necessary in the iterative updates of the estimated flaw image. We also discussed how to use the FFT to speed up the execution of the method and how to implement the method. The MEM approach was tested using both synthesized data, experimental data and inspection data, and the results were compared with those of the Wiener filter. Test results have shown that MEM has better resolution and noise reduction than the Wiener filter. Therefore, it can provide restoration results with a cleaner background and sharper edges. The major disadvantage of the MEM approach is that it is much more computationally intensive than the Wiener filter due to its iterative nature. On a DEC 5000 workstation, the Wiener filter took about 1-2 seconds of CPU time while the MEM approach took several minutes of CPU time.

CHAPTER 12. EDDY CURRENT IMAGE RESTORATION USING SIMULATED ANNEALING

Introduction

In the preceding two chapters we discussed linear eddy current image restoration techniques based on Wiener filtering and the maximum entropy method. Both linear restoration methods are based on an approximate linear forward model derived using the Born approximation. Therefore, the accuracy of the linear restoration methods are dependent on the validity of the Born approximation for the test geometry and the flaw configuration. From our test results we observed that although the linear restoration methods can provide reasonably good results for flaws that are much smaller than the coil, they generally break down when the flaw is larger than the coil or when the flaw has a high aspect ratio. To improve restoration results for these cases, we must study nonlinear image restoration methods which take the nonlinear nature of eddy current into consideration. The nonlinear image restoration methods must be based on forward models which can accurately model the nonlinear interaction between the flaw and the coil. In this chapter, we discuss a nonlinear eddy current image restoration method based on simulated annealing and the fast eddy current forward model discussed in Chapter 6. The neural network based fast forward model provides an efficient way to accurately evaluate a candidate flaw image. Based on the forward model, the nonlinear image restoration problem can be formulated as a combinatorial optimization problem which can be solved by using the simulated annealing algorithm. We

tested the simulated annealing based approach using both synthesized impedance change data and laboratory data. Test results have shown that the nonlinear image restoration method based on simulated annealing can generate restoration results with better resolution, better regularity, and a higher signal-to-noise ratio than the linear restoration methods, although it has higher computational complexity.

The Simulated Annealing Algorithm

Basic concept

Simulated annealing [85, 86] is one of the most widely used combinatorial optimization methods. Since a lot of problems in our world can be formulated as a combinatorial optimization problem, there have been numerous applications of the simulated annealing algorithm in the literature. Applications of the simulated annealing algorithm include image reconstruction and restoration [87, 88], solution of an inverse problem [89], binary code and sequence synthesis [90, 91], synthesis of unequally spaced arrays [92], finite impulse response (FIR) and infinite impulse response (IIR) filter design [93], and placement in the physical design of Very Large Scale Integration (VLSI) circuits [94]. A detailed discussion of the simulated annealing algorithm can be found in Reference [85].

Simulated annealing borrows its idea from the physical annealing process in which a metal piece is heated to high temperature and then cooled slowly so that the crystal structure in the metal can reach a configuration state with minimum energy. The basic idea of simulated annealing is instead of always going downhill on the cost surface as in gradient descent based methods (e.g., steepest descent, conjugate gradient, and the MEM algorithm

discussed in Chapter 11), try to go downhill most of the time. In other words, there is a nonzero probability of the acceptance of new states which increases the cost or the energy. Such probability is controlled by a temperature parameter, which becomes smaller and smaller towards the end of the annealing process. As a result of this principle, two properties of simulated annealing make it different from conventional iterative optimization algorithms [51].

1. The simulated annealing algorithm need not get stuck at a local minimum, since there is always a nonzero probability of transition out of a local minimum when the temperature is not zero.
2. Simulated annealing exhibits a *divide-and-conquer* feature that is adaptive in nature. In other words, gross features of the final state of the system are seen at higher temperatures, while fine details of the state appear at low temperatures.

Metropolis algorithm

The Metropolis algorithm, based on Monte Carlo techniques, is a method to simulate the evolution of a physical system in a heat reservoir to thermal equilibrium. In each state transition a new state is generated based on a perturbation to the current state, and the new state is accepted if it has a smaller cost than the current state. If the new state has a larger cost, it is accepted with a probability given by the following Metropolis function

$$p(T) = e^{-\Delta c/T}, \tag{12.1}$$

where T is the current temperature, and Δc is the cost difference. If the temperature is kept constant, after a large number of state transitions the system will reach its thermal

equilibrium. At thermal equilibrium, the probability that the system is in a state X with cost c_X is given by the Boltzmann distribution [51] of

$$P_X = \frac{1}{Z}\exp(-\frac{c_X}{T}), \tag{12.2}$$

where Z is the partition function, defined by

$$Z = \sum_Y \exp(-\frac{c_Y}{T}), \tag{12.3}$$

where the summation is taken over all states Y with cost c_Y at temperature T. From equation (12.2) we can see that at high temperatures the Boltzmann distribution exhibits a uniform preference for all states, while at very low temperature, only the states with minimum cost have a nonzero probability of occurrence.

Asymptotic convergence

The simulated annealing algorithm can be modeled using a Markov chain since the distribution of the new states generated from the current state is only dependent on the current state. A Markov chain is described in terms of transition probabilities which in simulated annealing are given by the Metropolis function. Since the Metropolis function is controlled by the temperature, the simulated annealing algorithm must be modeled using an inhomogeneous Markov chain when temperature changes. The asymptotic convergence to the global minimum of the simulated annealing algorithm has been proven by Geman and Geman [95]. However, in practice, due to limited execution time, only near-optimal solutions can be guaranteed. The trade-off between execution time and the optimality of solution is problem dependent and can only be decided experimentally.

Convergence speed

Since the simulated annealing algorithm is a very computationally intensive algorithm, there has been a lot of work in optimizing its convergence performance. Most of these works fall within the following three categories: careful perturbation [96], optimizing cooling schedule [97], and parallel simulated annealing [98]. Methods based on careful perturbation use range-limiting or heuristics in generating the new state so as to minimize the probability of rejection. Studies on optimizing a cooling schedule [97] indicate that a change in cooling schedule does not improve both the solution quality and execution by a large amount. More recently there have been studies on various parallel simulated annealing algorithms [98] that parallelize a single Markov chain or implement multiple Markov chains. It has been shown that these parallel implementations can significantly improve the convergence speed of simulated annealing, if a parallel computer is used.

Implementation of the Simulated Annealing Algorithm for the Eddy Current Image Restoration Problem

The nonlinear forward model

As discussed in Chapter 6, the nonlinear forward model is based on a radial basis function neural network. The neural network forward model learns the nonlinear mapping from a binary flaw image to a complex impedance change image from training samples obtained through numerical computations or experiments. To simplify the generation of the training data set, we require that there be only one flaw in the flaw image. With this restriction, we can form the training data set by using impedance change images of flaws with

various shapes and sizes. The reason to use the neural network based model instead of a numerical model is due to the dramatic speed improvement of the neural network model over the numerical models. Since in the simulated annealing algorithm a large number of forward computations must be performed, the speed of the forward model is critical to the practicality of the nonlinear image restoration method.

To generate the fast forward model, we first create a set of flaw images based on our prior information on flaw shape and size. For instance, if we know that the flaw is a crack, we can use rectangular flaws with various lengths and widths. If we have very little information about the flaw shape, we may need to create a training set that contains flaws with different shapes and sizes. The number of training flaws is dependent on the dimensionality of the solution space and the desired reconstruction error. Theoretical bounds on the approximation error and the guidelines for training data set selection can be found in Appendix B. After a set of training flaws is created, we can use a numerical model (e.g., the volume integral model with wavelet expansion discussed in Chapter 5) to compute the corresponding impedance change images. If there is no numerical model available for the test geometry, experiments can be made to obtain the training data set.

New state generation

The implementation of a new state generation step is critical to the convergence performance of the simulated annealing algorithm. In this step, a new state is obtained by applying a perturbation to the current state. The new state should be close enough to the current solution to preserve the stability of the convergence (range-limiting). Also, sufficient

randomness must be incorporated into the new state generation step so that the asymptotic convergence can be guaranteed. Despite the randomness requirement, applying some prior information to the solution in new state generation may improve convergence speed and the probability of generating a better solution.

In our problem, a solution is a binary image representing the shape of a flaw. Since we assume that there is only one flaw in the region of interest, in a valid solution all the white pixels representing the flaw should be connected together. In the new state generation step we must ensure that a valid new state is generated from a valid old state. This cannot be accomplished by using a simple random pixel flip scheme because it may generate an invalid state. To deal with this problem, we developed a new state generation algorithm in which a new state is obtained by randomly expanding or shrinking the old state. We have also incorporated a heuristic based on the energy of the impedance change image to guide the flaw growth along the correct direction. The algorithm is described as follows:

1. Randomly decide whether to add a pixel to or to remove a pixel from the current state. The probability of adding a pixel is P_{ar}, which is adaptively adjusted according to the energy of the impedance change image corresponding to the current state. The energy of a complex impedance change image is defined by

$$E = \sum_i \sum_j \left| \Delta Z_{ij} \right|^2 . \tag{12.4}$$

Let E_0 be the energy of the impedance change measurement and E_1 be the energy of the impedance change image corresponding to the current solution (which is computed in the cost evaluation step by using the forward model). Then if $E_0 > E_1$, set

$$P_{ar} = \begin{cases} P_{ar} + \Delta P, & \text{if } P_{ar} + \Delta P < 1 - P_{\min}, \\ 1 - P_{\min}, & \text{otherwise}, \end{cases} \tag{12.5}$$

and if $E_0 < E_1$, set

$$P_{ar} = \begin{cases} P_{ar} - \Delta P, & \text{if } P_{ar} - \Delta P > P_{\min}, \\ P_{\min}, & \text{otherwise}, \end{cases} \tag{12.6}$$

where ΔP is a small probability increment and $P_{\min}$ is the minimum probability for adding or removing a pixel. A typical value we used for ΔP is 0.05, for $P_{\min}$ is 0.2, and P_{ar} is usually initialized as 0.5. The purpose of the above algorithm is to create a new state which is likely to be close to the correct solution. If the current solution corresponds to an impedance change image with an energy smaller than desired, then it is very likely that the actual flaw is larger than the flaw in the current solution, and vice versa. In most cases, we can utilize this relationship between the size of the flaw and the energy of the impedance change image to create a new state that is closer to the optimal solution. However, this relationship is not always true due to the nonlinear nature of eddy current measurements. Due to this, we used a minimum probability of add/remove to ensure that the correct solution will not be missed. If $P_{\min}$ is not zero, the asymptotic convergence of the simulated annealing algorithm can still be guaranteed.

2. If a pixel needs to be added, the black boundary of the flaw (the set of pixels with value 0 that are neighbors of the white boundary (pixel value equal to 1) of the flaw region) in the current solution is generated using an 8-neighbor boundary search algorithm [78]. Among all the pixels on the black boundary, one is randomly selected and added to the flaw.

3. If a pixel needs to be deleted, the white boundary of the flaw in the current solution is generated using an 8-neighbor boundary search algorithm [78]. Among all the pixels on the white boundary, one is randomly selected and removed from the flaw. The selected pixel must not change the connectivity of the flaw (i.e., removing it will not let the flaw region be separated into two disconnected flaw regions). If this condition is not satisfied, the random selection is repeated until a pixel satisfying the condition is found and removed.

Cost evaluation

Cost evaluation is the process of evaluating how close a solution is to the ideal solution. It is the objective function of the simulated annealing algorithm. For our problem, we can use a simple cost function based on the mean square error between the predicted impedance change and the measured impedance change

$$C = \sum_i \sum_j \left| \Delta Z_{ij} - \Delta Z_{ij}^0 \right|^2, \tag{12.7}$$

where ΔZ_{ij} is the impedance change obtained from the current solution, and ΔZ_{ij}^0 is the measured impedance change. ΔZ_{ij} is obtained using the neural network fast forward model. To deal with the ill-posedness of the eddy current problem, in addition to the mean square error, we added a smoothing term to the cost function

$$C = \sum_i \sum_j \left| \Delta Z_{ij} - \Delta Z_{ij}^0 \right|^2 + \lambda\, m, \tag{12.8}$$

where λ is a regularization constant which controls the relative importance of the smoothing term and the mean square error, and m is the number of direction changes on the flaw boundary (number of corners of the flaw). The value of λ is related to the signal-to-noise ratio of the data. To obtain m, first we initialize m to be zero and then find the boundary of the flaw by using the 8-neighbor boundary tracking algorithm. Then starting from an arbitrary pixel on the boundary we check if its two neighbor boundary pixels are along the same direction. If they are not along the same direction, m is incremented by 1. After the current pixel is checked, we go to one of its neighbor boundary pixels and repeat the check. The process is finished when we return back to the starting pixel. Use of this smoothing term forces the solution to have a smooth boundary, or, in other words, to have a small number of corners. By changing the regularization constant, we can control the smoothness of the solution. The selection of the regularization constant is usually decided by the signal-to-noise ratio of the data. When the data has low SNR, one should use a large λ to reduce the effect of noise.

Cooling schedule

A cooling schedule consists of the initial high temperature, the end low temperature, the number of transitions per temperature, and a function for updating the temperature. A cooling schedule must make a trade-off between two conflicting requirements - fast speed and convergence to optimal or near-optimal results. The selection of the high temperature and the low temperature is related to the maximum and minimum values of the cost function which are generally not available before the annealing process. We used a heuristic based on

the energy of the measurement data to guide the selection of the high temperature and the low temperature. The energy of the measured impedance change is defined by

$$E_0 = \sum_i \sum_j \left|\Delta Z_{ij}^0\right|^2. \tag{12.9}$$

Comparing equations (12.8) and (12.9) we know that E_0 is equal to the cost when the candidate flaw image does not contain any flaw region (in this case $\Delta Z_{ij} = 0$ and $m = 0$). Therefore, E_0 gives a rough estimate of the maximum cost. Also, if noise in the impedance change measurement is not correlated to the ideal impedance change, then the minimum cost

$$C_{\min} \approx P_n = \frac{E_0}{SNR + 1}, \tag{12.10}$$

where P_n is the noise power and SNR is the signal-to-noise ratio of the data. We can then define the high temperature as

$$T_{high} = K_{high} C_{\max} = K_{high} E_0, \tag{12.11}$$

and define the low temperature as

$$T_{low} = K_{low} C_{\min} = K_{low} \frac{E_0}{SNR + 1}, \tag{12.12}$$

where K_{high} and K_{low} are constants. A typical value we used for K_{high} is 10 and for K_{low} is 0.1.

The rate of decrement in temperature is dependent on the regularity of the cost surface. If the temperature is decreasing too fast, we may lose the global optimum. Generally, the regularity of the cost surface is also not known, so we must experimentally decide the optimal rate of decrement. Since the exact form of the temperature decrement is not critical to the convergence performance, we use an exponential decrement function

$$T_{new} = \alpha T_{old}, \tag{12.13}$$

where $0 < \alpha < 1$ is a cooling constant which must be decided experimentally.

Summary of the algorithm

The simulated annealing based nonlinear eddy current image restoration method can be summarized as:

1. Select the training flaw set by using our prior information about flaw shape.
2. Use a numerical model or experimental measurements to obtain the impedance change images for the training flaw set.
3. Establish the fast forward model using the training data set and training algorithms discussed in Chapter 6.
4. Select the parameters for the cooling schedule as discussed above.
5. Generate a random initial flaw image. Since at high temperature simulated annealing searches over a large space, the selection of the initial state is not crucial to the performance of the algorithm.
6. From the current state, generate a new state using the adaptive growth algorithm. For the new state, calculate its corresponding impedance change image using the fast forward model and then compute its cost. Accept this new state if it has a lower cost than the current state. Otherwise, generate a random number that is uniformly distributed in 0 to 1 and compare it with the probability given by the Metropolis function. If the random number is smaller, accept this new state. Otherwise, repeat the above process until a transition occurs.

7. After a certain number of transitions, update the temperature using equation (12.13).
8. Stop if the temperature is smaller than the low temperature and return to the state with the smallest cost found in the annealing process.

Further considerations for laboratory data and inspection data

To apply the simulated annealing method to laboratory data and inspection data, we need to consider several practical issues. The first issue is how to build the training data set for the neural network forward model. If a numerical model is available for the test geometry, then we can use it to create a set of training samples. If a numerical model is not available, then we must make laboratory measurements on flaws with various sizes and shapes to create a training data set. The impedance change measurements used for this purpose must have high SNR to minimize modeling error. There are also two other problems associated with inspection data.

Alignment with forward model

The impedance change measurements obtained from field inspection must be aligned to be in position with the training data set so that a correct comparison can be made. This alignment problem is difficult because inspection data are usually noisy and distorted. A simple center-of-gravity computation may not give the correct center of the flaw in the impedance change image. To solve this problem, we used an approach which does not directly solve for the center of the actual flaw. In this approach, the position of the flaw is also one of the parameters to be optimized. In each cost evaluation step, we compute the center of the flaw region in the current solution by using the equations

$$x_c = \frac{\sum_{i,j} i \times f_{ij}}{\sum_{i,j} f_{ij}}, \tag{12.14}$$

and

$$y_c = \frac{\sum_{i,j} j \times f_{ij}}{\sum_{i,j} f_{ij}}, \tag{12.15}$$

where f_{ij} is the current flaw image. According to the center position obtained above, we shift the center of the flaw to the center of the image, which is the default center for the training data set. Then the forward model is used to find the impedance change image for the shifted flaw. We can obtain the impedance change image for the original flaw by simply shifting the result of the forward model in the opposite direction. This impedance change image is then compared with the impedance change measurement to compute the mean square error. This approach allows the annealing process to find the correct flaw position because the minimum cost can only be reached when the flaw is centered correctly.

Calibration

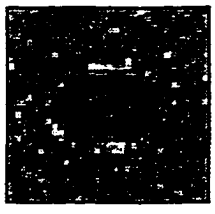
For inspection data, calibration is another problem that needs to be considered. If the data is collected using a probe that is of the same type but different from the probe used to obtain the training data set, the impedance change of the inspection data may have different phase and amplitude than the training sample set due to variations in probe characteristics. Therefore, such inspection data must be calibrated before we can use the simulated annealing method. Calibration of the data can be accomplished by using the probe to collect data for a standard calibration tube and then comparing the data with calibration data obtained using the

probe that generates the training data set. According to differences in the two tubes' responses to some artificial flaws on the calibration tube, we can derive a calibration constant which is then multiplied with the inspection data.

Test Results

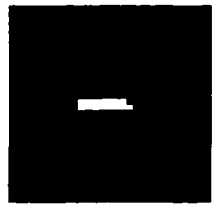
Test results for synthesized data

We have tested our simulated annealing restoration method using synthesized impedance change data. The test was based on synthesized impedance change images for a EDM slot with different SNR obtained using a volume integral method forward model. The neural network fast forward model was trained by using the training data set shown in Fig. 6.3(a). The test data set consists of synthesized impedance-change data with signal-to-noise ratios of 20 dB, 10 dB and 0 dB. The real and imaginary parts of the impedance change, and the restoration for the 20 dB case are shown in Fig. 12.1. A smoothing constant (λ in equation (12.8)) of 1 was used in this case. The real and imaginary parts of the impedance change, and the restoration result for the 10 dB case are shown in Fig. 12.2. A smoothing constant of 5 was



(a) Re(Z)

(b) Im(Z)

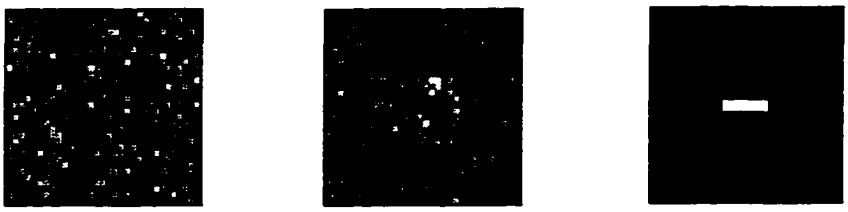


(c) Restoration result

Fig. 12.1. Test results for the 20 dB data.

(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 12.2. Test results for the 10 dB data.



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 12.3. Test results for the 0 dB data.

used in this case. The real and imaginary parts of the impedance change, and the restoration result for the 0 dB case are shown in Fig. 12.3. A smoothing constant of 10 was used in this case.

As can be seen from the results shown above, in the 10 dB and the 0 dB case, the restoration result is identical to the ideal flaw image. For all test results presented above, the initial high temperature (equation (12.11)) was 1000, the final low temperature (equation (12.12)) was 0.1, the cooling constant (equation (12.13)) was 0.8, the number of transitions per temperature was 10, ΔP was 0.05, and $P_{\min}$ was 0.2. Fig. 12.4 illustrates the change of the cost as a function of the number of state transitions in a typical test of the simulated annealing based nonlinear image

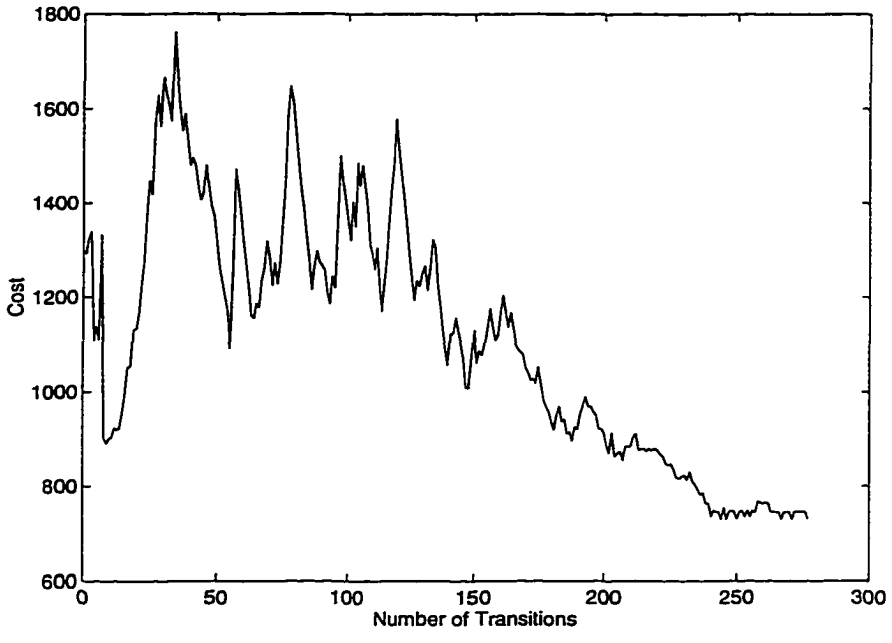


Fig. 12.4. The change in cost as a function of the number of state transitions.

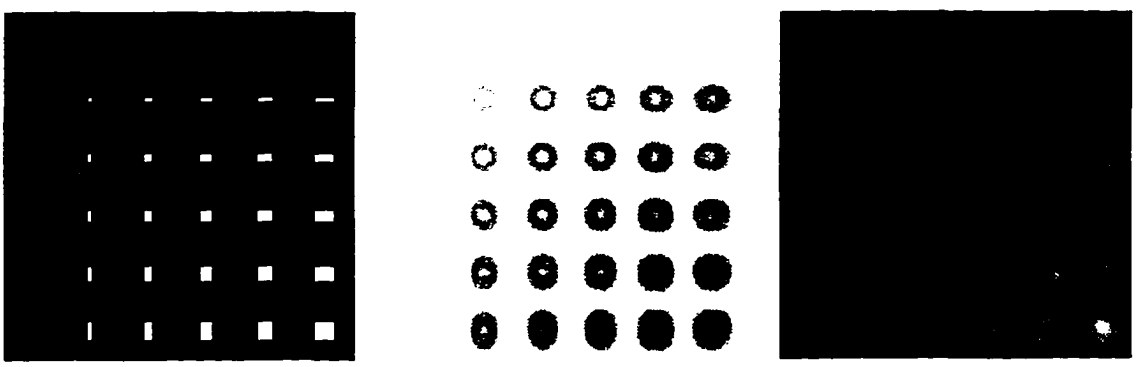
restoration method. As can be seen from Fig. 12.4, when the temperature is high (the first part of the curve), the change in cost is less regular and there are many transitions to states with higher cost. As the temperature decreases to a certain point, the cost tends to decrease monotonically. The cost becomes almost a constant when the temperature is sufficiently low. At this time, the simulated annealing algorithm has reached the optimal solution or a near-optimal solution. Notice that the number of states evaluated in the simulated annealing algorithm is much larger than the number of state transitions that occurred. This is due to the very small probability of accepting a new state with higher cost at a low temperature. On average, a state transition occurring at a low temperature requires a large number of cost evaluations, or in other words, a large number of forward computations. Therefore, the selection of the low temperature is more difficult than the selection of the high temperature

and a tradeoff must be made between the convergence to good solution and the convergence speed.

Test results for laboratory data

To evaluate the performance of the simulated annealing method on experimental data, we tested it using laboratory data derived from a two-dimensional impedance change measurement for a surface crack on a thick aluminum plate. The crack had a length of 4.04 mm, a width of 0.21 mm, and a depth of 0.916 mm. The aluminum plate was thick enough to be considered as a half space. The coil used had an inner radius of 3.8 mm, an outer radius of 5.63 mm, and a height of 2.64 mm. A two-dimensional scan was made on a 64 by 64 grid, with a grid separation of 0.4 mm. Before restoration, the data were preprocessed using polynomial background removal and spline smoothing.

To generate the training data set for the neural network forward model, we created 36 artificial flaws with various lengths and widths. The impedance change images for those



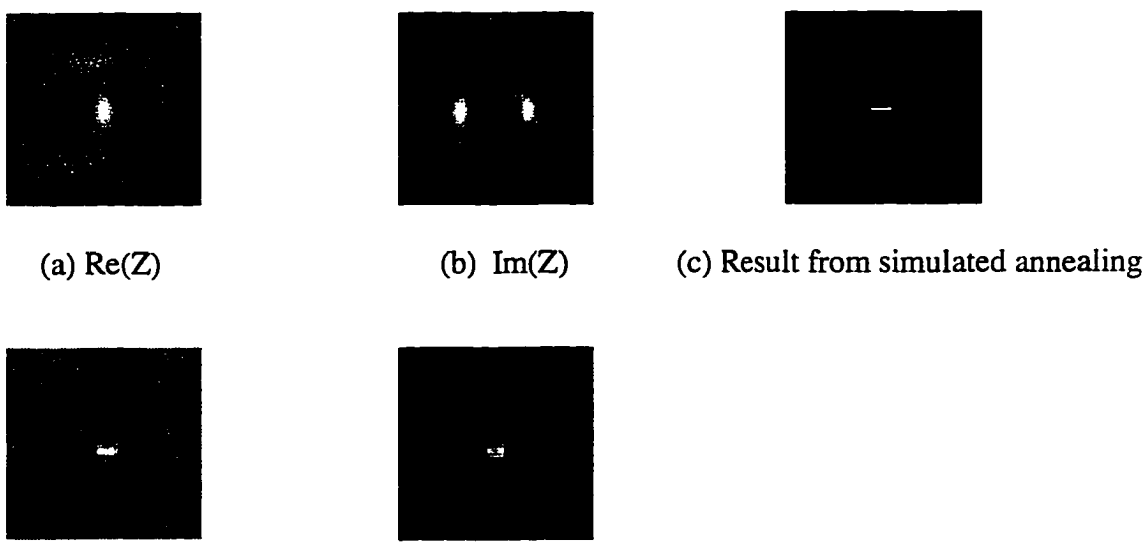
(a) Flaw images (b) Re(Z) images (c) Im(Z) images

Fig. 12.5. The training data set for creating the neural network forward model.

artificial flaws were then computed by using the volume integral model. The training flaw images and the real and imaginary impedance change images are shown in Fig. 12.5.

After the training data set was created, we generated a fast forward model by using the radial basis function neural network model. Then we applied the simulated annealing method to the laboratory data using this forward model. For this test, the initial high temperature was 10000, the final low temperature was 5, the cooling constant was 0.8, the number of transitions per temperature was 10, ΔP was 0.05, P_{min} was 0.2, and the smoothing constant was 20.

The real part and imaginary part impedance change images of the laboratory data and the restoration results from simulated annealing are shown in Fig. 12.6. As a comparison, Fig. 12.6 also gives the restoration results from Wiener filtering and the maximum entropy



(a) Re(Z) (b) Im(Z) (c) Result from simulated annealing

(d) Result of Wiener filtering (e) Result of MEM

Fig. 12.6. Test results from the laboratory data.

method. In the restoration results from all the three methods, the restored flaw is shorter than the actual flaw. This is thought to be a result of the background noise in the laboratory data and the distortion introduced to the data during preprocessing. From Fig. 12.6 we can see that the restoration result from the simulated annealing method is better than that of Wiener filtering and MEM in term of resolution and signal-to-noise ratio. Another advantage of the simulated annealing method is that in its restoration result there is no distortion due to a linear approximation, which can usually be found in the results of the two linear methods for relatively long cracks. The major disadvantage of the simulated annealing approach, however, is its high computational complexity. A typical execution of the simulated annealing image restoration requires performing several thousands of forward computations, which takes about 20-30 minutes on a DEC 5000 workstation. Apparently this speed is not practical for most real-world applications.

CHAPTER 13. EDDY CURRENT IMAGE RESTORATION USING GENETIC ALGORITHMS

Introduction

In the last chapter we discussed a nonlinear eddy current image restoration technique based on the simulated annealing algorithm. Test results showed that the simulated annealing based approach has good potential for achieving high quality restoration results compared with the linear restoration methods. However, test results have also revealed that the method is computationally slow for practical applications. The convergence of this method to a near-optimal solution usually requires several thousand forward computations which consumes a large amount of computer time even when the neural network based fast forward model is used. To improve the speed performance of the nonlinear image restoration method, we need to develop a method that is more efficient in the search for a near-optimal result. In this chapter we discuss a nonlinear image restoration method based on genetic algorithms. More detail on this method can be found in Appendix B. Methods based upon genetic algorithms have been shown to have faster convergence speed than that of the simulated annealing algorithms in many applications [99-109]. We tested the genetic algorithm based approach using both synthesized impedance change data and laboratory data. We also compared the test results with the results of the simulated annealing based approach. Our test results showed that the genetic algorithm based approach can generate a near-optimal result in a

much shorter computational time than the simulated annealing based approach. In all the tests we made the speed improvement was larger than a factor of ten.

Background of Genetic Algorithms

Basic concept

Like the simulated annealing algorithm, genetic algorithms are used for solving combinatorial optimization problems. Genetic algorithms were developed originally in the field of artificial intelligence by John Holland [101] more than 20 years ago. But only recently has it been widely used in solving the combinatorial optimization problem. Applications of genetic algorithms include configuration of radial basis function networks [102], image reconstruction [105], design and analysis of control systems [106], solution of electromagnetic inverse problems [107], optimal design of digital filters [108], and image segmentation [109].

Genetic algorithms simulate the natural evolution process. More specifically, they involve the following steps:

1. Encode the candidate solutions of a problem in binary strings or arrays of numbers. This encoded representation is called a chromosome.
2. Define a fitness function for the candidate solutions. For an optimization problem, the fitness function is related to the objective function to be optimized.
3. Define a selection process to generate an intermediate population from an initial population based on the fitness of each individual.

4. Define a mutation operation which brings random changes to the chromosome of each individual.
5. Define a crossover operation to exchange the genetic material of two individuals.
6. Perform an evolution process which is based on the above defined operations.

In many applications, genetic algorithms have been found to be more efficient than the simulated annealing algorithm. This is mainly due to the crossover operation used to create new candidate solutions from the current population. The crossover operation can extend the region of search to be much larger than the region of search for the random perturbation used in simulated annealing. This increases the probability of finding a good candidate and improves convergence performance.

Schema theorem

Holland [101] has derived a formula for the average number of instances for a set of individuals based on the average fitness of the individuals. The formula, called schema theorem, applies to binary string encoded genetic algorithms with a one-point crossover operation and bit-flip mutation operation. A schema is a binary string with some don't-care bits. There may be many different binary strings belonging to the same schema. The schema theorem can be stated as

$$N(\mathbf{h},t+1) \geq N(\mathbf{h},t)\frac{f(\mathbf{h},t)}{\bar{f}(t)}\left[1-P_c\frac{\delta(\mathbf{h})}{l-1}-P_m o(\mathbf{h})\right], \tag{13.1}$$

where $N(\mathbf{h},t)$ is the expected number of instances of schema $\mathbf{h}$ in generation t, P_c is the crossover probability, P_m is the mutation probability, $f(\mathbf{h},t)$ is the average fitness value of

schema $\mathbf{h}$ in generation t, $\bar{f}(t)$ is the average fitness value of the population in generation t, l is the length of the binary string, $\delta(\mathbf{h})$ is the defining length of $\mathbf{h}$ (the distance between the outermost fixed positions), and $o(\mathbf{h})$ is the order of $\mathbf{h}$ (the number of fixed positions).

The schema theorem has some important implications. First, it means that the algorithm will converge to a uniform population of the best individual if no crossover and mutation are performed. Therefore, we should incorporate a sufficient amount of crossovers and mutations to create new individuals which might have higher fitness level. Second, because the expected number of instances is related to a schema, not a string, genetic algorithms actually implement an implicit parallelism because a string simultaneously represents 2^l different schemata (for each bit of the string, the schema can have either the same value as the string or a don't-care bit).

For applications where the assumptions (binary string encoding, one point crossover, simple bit-wise mutation) of the schema theorem are not satisfied, the general principle of the schema theorem can still apply. However, the definition of the crossover operation and the mutation operation should follow the concepts of the basic binary genetic algorithm.

Implementation of Genetic Algorithm Based Eddy Current Image Restoration Method

Encoding of the solution

The individuals in the genetic algorithm are the candidate solutions of the optimization problem. The first step of the genetic algorithm is to encode the candidate

solutions in a convenient form so that the mutation and crossover operators can be efficiently defined. In our image restoration problem, the candidate solutions are images containing possible flaw shapes. According to the neural network forward model we developed, the flaw image is defined as a binary image for the conductivity distribution in the flaw region. The binary flaw image can also be considered as a binary string if we map the two-dimensional array of binary numbers into a one-dimensional array of binary numbers by rearranging the indices. The candidate solution can then be expressed as

$$\mathbf{X} = [x_1, x_2, \cdots, x_P], x_i = \{0, 1\}, i = 1, \cdots, P, \tag{13.2}$$

where $P = M \times N$ for a M by N image. Because in the fast forward model we have assumed that there is only one flaw region in the flaw image, not every binary string given by equation (13.2) is a valid candidate solution. This adds extra complexity to the definitions of the mutation operation and the crossover operation as we need to make sure the results of these operations are also valid solutions.

Fitness function

The fitness function defines how well an individual is able to survive in the evolution. In an optimization problem, the fitness function defines how close a solution is to the optimal solution. In our image restoration problem, we can define the fitness function as the reciprocal of the cost function we used for the simulated annealing algorithm as

$$f(\mathbf{X}) = 1 / C(\mathbf{X}), \tag{13.3}$$

where $\mathbf{X}$ is the binary string representation of the candidate flaw image, and the cost function is given by

$$C(\mathbf{X}) = \sum_i \sum_j \left|\Delta Z_{ij} - \Delta Z_{ij}^0\right|^2 + \lambda m, \tag{13.4}$$

where λ is a regularization constant which controls the relative importance of the smoothing term and the mean square error, and m is the number of directional changes on the flaw boundary (number of corners in the flaw). Details of equation (13.4) can be found in Chapter 12.

Population selection

The most widely used selection scheme is the proportionate selection scheme in which the number of offspring for a given candidate is proportional to its fitness. However, because of the noise in the impedance change measurement, the minimum cost has a nonzero value, equation (12.10). Also, because the fitness defined in equation (13.2) introduces a nonlinear scaling of the cost function, the fitness difference between a good solution and a bad solution can be very small compared to the base value of the fitness. If the proportionate selection scheme is used, the number of offspring generated by the good solution is only slightly larger than that of the bad solution. This means that the selection process has a very small impact on the fitness distribution of the intermediate population. As a result of this, the convergence of the genetic algorithm is very slow. To avoid this situation, we used the rank-based selection scheme [99]. In the rank-based selection, the expected number of offspring for a candidate solution is dependent on its fitness rank, not the absolute value of its fitness. Therefore, rank-based selection can guarantee that the good solution has much more offspring than the bad solution.

The rank-based selection is implemented as follows:

1. Define a probability P_r for the rank-based selection.
2. Sort the candidate solutions according to a descending order of fitness.
3. Start with the candidate solution with the highest fitness.
4. Generate a uniformly distributed random number between 0 to 1. If this number is smaller than P_r, choose the current candidate, otherwise go to the next candidate and repeat the process.
5. Repeat step 3 and step 4 until all individuals in the intermediate population are generated.

Crossover

The crossover operation is a very important operation in genetic algorithms. The crossover operation should randomly incorporate the genetic materials of the two parents to generate offspring without any preference towards either one of the parents. In an optimization problem with solutions encoded as short binary strings, the crossover operation can be simply defined as exchanging part of the binary strings between the two parents. The crossover operation in our problem is harder to define, because not all binary images are valid flaw images. In our forward model we require that there be only one single flaw region in the flaw image. If we simply exchange part of two binary images for two parents, we may end up with an invalid flaw image which contains two or more unconnected flaw regions. For this problem, we must define a crossover operation that generates a valid flaw image from two valid flaw images with a different flaw location and flaw shape. We developed an algorithm for the crossover operation that is based on examining the intersection and union of

the flaw regions of the two parents. The intersection of the two flaw regions contains genetic material common to the two parents, which is likely to be an important part of the optimal solution. Therefore, we initialize the offspring with the intersection and then randomly grow it in the union of the two flaw regions. By letting the parents be $\mathbf{X}_1$ and $\mathbf{X}_2$, $\mathbf{I} = \mathbf{X}_1 \cap \mathbf{X}_2$, and $\mathbf{U} = \mathbf{X}_1 \cup \mathbf{X}_2$, we can use the following algorithm to create an offspring $\mathbf{Y}$.

1. If $\mathbf{I} = \varnothing$, $\mathbf{Y} = \mathbf{X}_1$ with probability 0.5 and $\mathbf{Y} = \mathbf{X}_2$ with probability 0.5. In this case, we do not need to perform a crossover between the two parents because at least one of them is not a good solution. The reason for this is that the optimal solution must have both the correct flaw shape and flaw location since the two parents have different flaw locations. We can conclude that at least one of them is not a good solution so that it will be soon eliminated from the population.
2. If $\mathbf{I}$ has only one connected flaw region, then the offspring are generated in the following way:

   (1) Set $\mathbf{Y} = \mathbf{I}$.

   (2) While $A(\mathbf{Y}) < \frac{1}{2}\left(A(\mathbf{X}_1) + A(\mathbf{X}_2)\right)$, where $A(\mathbf{X})$ is a function which returns the flaw area of a flaw image $\mathbf{X}$,

   a. Let $\mathbf{B}$ be a binary image containing the white boundary (pixel value equal to 0) of $\mathbf{Y}$, $\mathbf{B}_1 = \mathbf{B} \cap \mathbf{X}_1$, and $\mathbf{B}_2 = \mathbf{B} \cap \mathbf{X}_2$.

   b. $\mathbf{Y} = \mathbf{Y} \cup \mathbf{B}_1$ with probability 0.5 and $\mathbf{Y} = \mathbf{Y} \cup \mathbf{B}_2$ with probability 0.5.

   In this case, the offspring is first initialized using the intersection region of the flaw regions of the two parents, and then it randomly grows in the union of the flaw regions

until it reaches a reasonable size (in this case, the average size of the two flaw regions in the two parents).

3. If $\mathbf{I}$ has more than one flaw region, $\mathbf{Y} = \mathbf{X}_1$ with probability 0.5 and $\mathbf{Y} = \mathbf{X}_2$ with probability 0.5. In this case, a meaningful and valid crossover operation is very hard to define due to the complexity of possible combinations of intersection regions. Since this case is rare, we can ignore the crossover operation for this case and let the mutation operation take over. As a result of the mutation operation, even if the two individuals are paired off again in later generations, a case 1 or case 2 situation will finally be reached so that there is no danger of deadlock.

Mutation

The mutation operation in the genetic algorithm is very similar to the new state generation step in the simulated annealing algorithm. It also applies a random perturbation to a candidate solution. The implementation of the mutation operation is also based on the random expanding or shrinking of the flaw region in the flaw image. The only difference is that now the accumulative update of P_{ar} (12.5, 12.6) is no longer used because there are multiple candidate solutions in the population. Instead, P_{ar} is decided on a per individual basis as

$$P_{ar} = \begin{cases} P_{\min}, & \text{if } E_1 > E_0, \\ 1 - P_{\min}, & \text{otherwise}, \end{cases} \tag{13.5}$$

where E_0 is the energy of the impedance change measurement and E_1 is the energy of the impedance change image corresponding to the current solution. Details on the mutation operation can be found in the descriptions of the new state generation step in Chapter 12.

Summary of the algorithm

1. Select the training flaw set by using prior information about the flaw shape.
2. Use a numerical model or experimental measurements to obtain the impedance change images for the training flaw set.
3. Establish the fast forward model using the training data set and training algorithms discussed in Chapter 6.
4. Select the crossover probability P_c, the mutation probability P_m, the probability for the rank based selection P_r, the population size N, and the maximum number of population K.
5. Initialize a population of N candidate flaw images using random guesses or results obtained from linear eddy current image restoration methods.
6. For each generation:
   (1) Evaluate the fitness of each individual.
   (2) Create an intermediate population using the rank based selection with the probability P_r.
   (3) Randomly pair off individuals and perform crossover with the probability P_c.
   (4) Perform mutations to individuals with the probability P_m.

3. Stop if the evolution reaches steady state or after K generations. Return the best individual found in the evolution process.

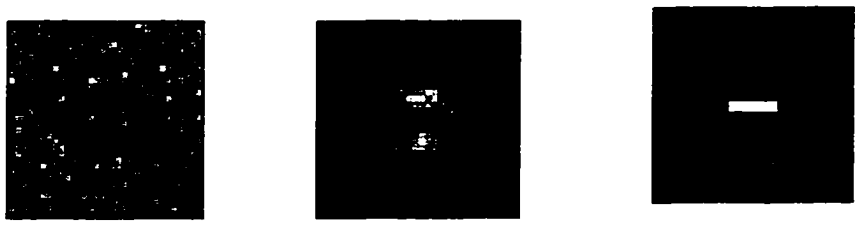
Test Results

Test results for synthesized data

We tested the genetic algorithm based approach using the set of 20, 10 and 0 dB synthesized impedance change data which we used to test the simulated annealing based approach. All common parameters used in this test were the same as that used in the test of the simulated annealing algorithm. The real and imaginary parts of the impedance change and the restoration result for the 20 dB case are shown in Fig. 13.1. A smoothing constant of 1 was used in this case. The real and imaginary parts of the impedance change and the restoration result for the 10 dB case are shown in Fig. 13.2. A smoothing constant of 5 was used in this case. The real and imaginary parts of the impedance change and the restoration

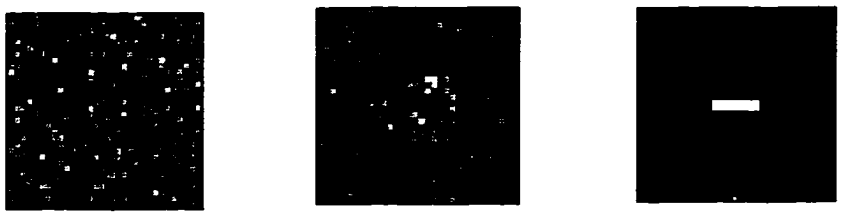
(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 13.1. Test result for the 20 dB data.



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 13.2. Test result for the 10 dB data.



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 13.3. Test result for the 0 dB data.

result for the 0 dB case is shown in Fig. 13.3. A smoothing constant of 10 was used in this case.

The above results were obtained with a crossover probability of 0.9, a mutation probability of 0.7, a probability for rank-based selection of 0.3, a population size of 10, and a maximum number of generations of 10. The initial population for the above tests was generated using rectangles with random widths and lengths. As can be seen from the above results, for the 10 dB case and the 0 dB case the restoration result is identical to the ideal flaw image, the same as the results of the simulated annealing based approach. Fig. 13.4 illustrates the maximum fitness in the population as a function of the number of generations

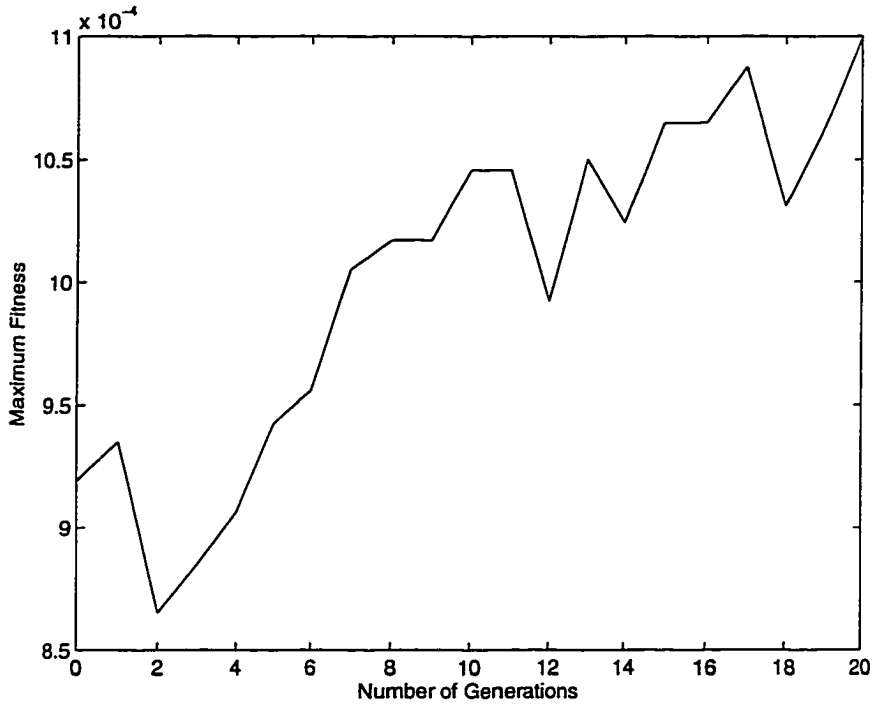


Fig. 13.4. Maximum fitness as a function of the number of generations.

for a typical execution of the algorithm. As can be seen from Fig. 13.4, the maximum fitness increases almost monotonically in the evolution process. Comparing Fig. 13.4 with Fig. 12.4, we can conclude that the genetic algorithm based approach could be more efficient in term of convergence speed if proper parameters and initial population are used. For all the test results shown above, the genetic algorithm based approach reached a near-optimal solution within 10 generations. Therefore, the above tests requires about 100 forward computations each, compared with 2000 - 3000 forward computations in the simulated annealing algorithm. We also observed the following properties of the algorithm from the tests:

1. The diversity of the initial population is very important to the convergence speed. Higher diversity leads to faster convergence. Since the nature of the crossover operation is as an

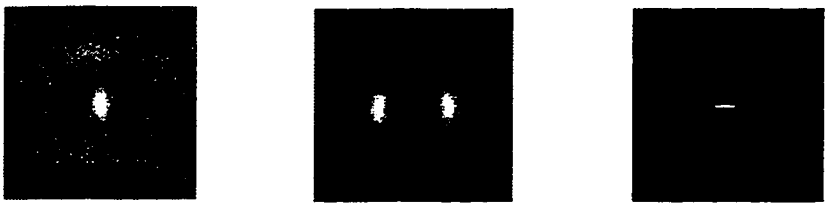
interpolation between the two parents, it is useful only when the two parents are not very similar. Therefore, if the initial population contains similar individuals, only the mutation operation is initially useful. Due to the local search nature of the mutation operation, it takes a long time for the algorithm to converge to a near-optimal solution. If the individuals are scattered randomly in the solution space, then the search by the genetic algorithm is more effective and a faster convergence speed can be achieved.

2. High crossover probability and high mutation probability leads to fast convergence. Although this is not always true for all problems, in our application we observed that high crossover probability and high mutation probability increase search activity within a fixed number of generations; thus, increasing the probability of finding a good solution.
3. The probability for the rank-based selection should be set so that the bottom half of the population is rarely selected. This is used to ensure that most individuals in the intermediate population have high fitness.

Test results for laboratory data

To evaluate the performance of the genetic algorithm based method on experimental data, we tested it using the same laboratory data used to test the simulated annealing algorithm. Details of the laboratory data and the training of the neural network fast model can be found in Chapter 12. The test was performed with a crossover probability of 0.9, a mutation probability of 0.7, a probability for rank-based selection of 0.3, a population size of 10, and a maximum number of generations of 10. The real part and imaginary part impedance change images of the laboratory data and the restoration result from the genetic

algorithm are shown in Fig. 13.5. From Fig. 13.5 we can see that the restoration result from the genetic algorithm is the same as that from the simulated annealing method. However, the execution time of the genetic algorithm based approach is less than one-tenth of the execution time of the simulated annealing based approach. The speed performance of the genetic algorithm based approach can be further improved if we use restoration results from linear restoration methods in the initial population. Considering the rapid progress in computer performance, we believe that the genetic algorithm based approach will become a practical solution of the eddy current image restoration problem in the near future.



(a) Re(Z) (b) Im(Z) (c) Restoration result

Fig. 13.5. Test results from the laboratory data.

CHAPTER 14. CONCLUSIONS

Summary of Major Results

Eddy current forward modeling

In this study we have developed an efficient implementation of the volume integral method for half space geometry and thin plate geometry. Based on careful considerations of numerical efficiency and stability, especially the stability of the integrations for the singular element, this implementation provides a solid basis for our understanding of the forward problem and the development of other forward models. To improve the speed of the volume integral method, we developed a wavelet expansion method which transforms the system matrix to a very sparse matrix. The wavelet expansion method can significantly reduce the computation involved in the volume integral method with only a small loss in accuracy. To further improve the speed of the forward model for applications in which the speed of the forward computation is critical, we also developed fast forward models based on artificial neural networks, especially the radial basis function neural networks. After being trained using training samples obtained from experiments or numerical modeling, the neural network based forward model can closely approximate an ideal numerical model in a subspace of the whole modeling space at an extremely high speed. The major disadvantage of the neural network forward model is that the training data set must be carefully selected so that the targeted subspace can be fully covered and the average approximation error is minimized.

The above mentioned three forward models can be selectively used to handle different modeling problem. The direct implementation of the volume integral method is the most accurate model, but it is also the most time consuming model. It should be used when the flaw volume is small so that the number of volume elements is limited. The volume integral method with wavelet expansion can be used to model relatively large flaws when the number of volume integral elements required makes the direct implementation method impractical. The neural network based forward models are used for applications where the modeling speed is the most important thing, and a certain amount of modeling error can be tolerated. Such applications include forward models used to solve the nonlinear inverse problem and forward models used for educational purposes.

Signal processing and flaw detection

In this study we developed signal processing techniques to enhance flaw signals and a detection method to find circumferential cracks. The set of signal processing techniques includes preprocessing techniques to remove background signals from liftoff and changes in tube diameter and to adjust the phase of the data so as to enhance signals from OD flaws. We also developed a background removal method based on principal component analysis to suppress signals from dents and scratches on the tube. The circumferential crack detection method is based on the shape characteristics of a circumferential crack signal and grayscale mathematical morphological operations.

The above mentioned processing techniques are usually used in series combination. Firstly, the inspection data is processed by preprocessing techniques to remove the null offset

and to rotate the OD flaw signals to the vertical channel. Then the vertical channel is processed by the background removal method based on principal component analysis to further reduce signals from other unwanted effects. Finally, the set of morphological operations are performed and a binary crack map is generated. Once the possible flaw indications are located, we can use image restoration methods to recover the correct flaw shape so that better flaw characterization can be achieved.

Eddy current image restoration methods

Another major part of this study was the image restoration methods to restore the flaw image from the blurred impedance change image. We developed two linear image restoration methods - the Wiener filter method and the maximum entropy method, and two nonlinear image restoration methods - the simulated annealing based method and the genetic algorithm based approach. The two linear image restoration methods are based on an approximate linear forward model derived using the reciprocity impedance change formula and the Born approximation. The Wiener filter method is a straightforward method. It can be easily applied when the point spread function is available. It involves a small amount of computation and generates less than ideal results. The maximum entropy method is an iterative method with higher computational complexity than the Wiener filter method. However, it generally provides restoration results with higher signal-to-noise ratio and resolution. Due to the error introduced in the linear approximation, both linear restoration methods may generate erratic results in some cases. To remedy this problem, the nonlinear image restoration methods are based on the neural network forward model which is more

accurate than the approximate linear forward model. Although the nonlinear image restoration methods are more time consuming than the linear methods, they can generate better restoration results and are reliable under various flaw and coil configurations. Generally the genetic algorithm based approach has much faster convergence than the simulated annealing based approach.

Whether to use the linear restoration methods or to use the nonlinear restoration methods in a specific application depends on the tradeoff between the quality of the restoration results and the speed of the restoration. Usually linear restoration methods can be used as a first-cut to process a large amount of data in a short time. For flaws that are of critical size, the nonlinear restoration methods can then be used to generate more accurate results.

Future Research

Although in this dissertation we presented a comprehensive study on many aspects of the processing and analysis of two-dimensional eddy current inspection data, there are still many problems to be investigated so that we can improve the performance of our processing techniques for practical applications. Two of these problems have special importance for industrial applications. The first problem is how to develop numerical forward models that accurately model tube geometry and industrial standard coils. The second problem is how to improve the speed of the nonlinear image restoration methods developed in this study so that they can be used in practical data analysis.

The numerical models currently available in the study of eddy current testing, including the volume integral method adopted in this study, usually simplify the modeling problem of tube inspection with thin plate geometry and an air-core coil. Although this simplification is valid in some cases, we would like to have a more accurate model that takes the tube geometry and the non-ideal nature of industrial standard coils into consideration. In industrial standard coils ferrite cores are usually used to increase the magnitude of the coil response. Due to the nonlinear nature of the ferrite core, the incident field of the coil cannot be analytically computed. Instead, complex numerical methods are necessary to compute the incident fields of industrial standard coils. The tube geometry brings another complication - because the lack of symmetry in this case, there is by far no analytical solutions for the incident field of a pancake coil placed above the tube internal surface and for the Green's function in tube geometry. Therefore, the volume integral method cannot be used. The finite element method may be able to model the complete tube geometry. However, the complexity of the finite element method prevents it from being widely used in industrial practice. It can be predicted that advances in this practical modeling problem can lead to better understanding of the tubing inspection problem and improvements in processing techniques for tubing inspection data.

To increase the practicality of the nonlinear image processing restoration techniques, we need to further improve their speed performance. Although it is believed that the rapid progress in computer performance will ultimately fill up the speed gap, in the near future we still have to study methods for reducing the amount of computation involved in the nonlinear restoration. Two approaches are possible: (1) improve the speed performance of the neural

network forward model, and (2) reduce the number of forward computations required to reach a near-optimal solution. To use the first approach, we need to reduce the size of the radial basis function (RBF) neural network since the complexity of the forward computation for an RBF network is dependent on the number of nodes in each layer. To reduce the number of nodes in the input layer and output layer, we can compress the input and the output as discussed in Chapter 6. To reduce the number of nodes in the hidden layer, we need to study better clustering methods so that the approximation capability of the radial basis function neural network can be preserved even when the number of cluster centers is small. To use the second approach, we need to further investigate the convergence properties of the combinatorial optimization algorithms, especially the genetic algorithm. Since the performance of the genetic algorithm is dependent on a number of parameters used in the algorithm, and because there are many variations in the genetic algorithm, this investigation requires a large amount of testing and performance tuning. However, once the above two approaches have been carefully studied, we should be able to significantly improve the speed performance of the genetic algorithm based nonlinear image restoration method.

APPENDIX A

WAVELETS IN THE SOLUTION OF THE VOLUME INTEGRAL EQUATION: APPLICATION TO EDDY CURRENT MODELING

A paper accepted by
Journal of Applied Physics

Bing Wang[a)]
Department of Electrical and Computer Engineering
and Center for NDE
Iowa State University
Ames, IA 50011

John C. Moulder[b)]
Center for NDE
Iowa State University
Ames, IA 50011

John P. Basart[c)]
Department of Electrical and Computer Engineering
and Center for NDE
Iowa State University
Ames, IA 50011

[a)] Electronic mail: bingwang@iastate.edu
[b)] Electronic mail: jmoulder@cnde.iastate.edu
[c)] Electronic mail: jpbasart@iastate.edu

Abstract

There is growing interest in the applications of wavelets as basis functions in solutions of integral equations, especially in the area of electromagnetic field problems. In this paper we apply a wavelet expansion to the solution of the three-dimensional eddy current modeling problem based on the volume integral method. Although this method shows promise for eddy current modeling of three-dimensional flaws, it is restricted by the computing power required to solve a large linear system. In this paper we show that applying a wavelet basis to the volume integral method can dramatically reduce the size of the linear system to be solved. In our approach, the unknown total field is expressed as a twofold summation of shifted and dilated forms of a properly chosen basis function, which is often referred to as the mother wavelet. The wavelet expansion can adaptively fit itself to the total field distribution by distributing the localized functions near the flaw boundary, where the field change is large, and the more spatially diffused functions over the interior of the flaw where the total field tends to be smooth. The approach is thus best suited to modeling large three-dimensional flaws where the large number of elements used in the volume integral method requires extremely large memory space and computational capacity. The feasibility of the wavelet method is discussed in the context of the physical nature of eddy current modeling problems. Numerical examples using both Haar wavelets and Daubechies compactly supported wavelets with periodic extension are given. The results of the wavelet method are also compared with experimental results from a cylindrical flat-bottom hole in an aluminum plate. These numerical examples and comparisons indicate that the wavelet method can greatly reduce the numerical complexity of the problem with negligible loss in accuracy.

1. Introduction

Eddy current testing is a widely used nondestructive evaluation (NDE) technique in which flaw information is extracted from the impedance change of a coil placed above a metal testpiece. Two common applications of eddy current NDE are the inspection of heat exchanger tubes in steam generators of nuclear power plants and detection of hidden corrosion in the lap-splices of aircraft skins. To obtain quantitative information about flaw size and shape, we would like to have a theoretical model which is able to predict the impedance change of a coil for different flaws in the test geometry. Analytical solutions exist for simple test geometry and flaws with good symmetry properties. However, for flaws with irregular shapes in complex geometry, an analytical solution usually is not available so we must find a numerical solution. There have been several numerical models in the literature, e.g., the finite element method[1] , the boundary element method[2] , the volume integral method[3,4,5] and methods based on variational formulas.[6] The finite element method is a general method which can be used in various flaw and geometry settings. However, it is very complex and inherently computationally intensive. The boundary element method reduces the computation by using only elements on the boundary of the flaw, but it cannot be used easily to model irregular flaws. Variational methods can also reduce the computation time, but their convergence is not guaranteed. The volume integral method has shown good potential due to its capability of modeling a three dimensional flaw with arbitrary shape. It is also straightforward to apply the volume integral method if the Green's function of the given geometry is known. A limitation of the volume integral method is that it requires a large number of volume elements to accurately model a large three-dimensional flaw. This usually results in the inability of the volume integral method to model large flaws with conventional computing resources, such as PCs and workstations.

To solve an integral equation numerically, variational formulas are used to discretize the integral equation and change it into linear equations by using a set of basis functions. It is well known that the selection of basis functions in the solution of integral equations plays an

important role in the sparsity of the resultant system matrix. Recently there has been growing research interest in the application of wavelets as basis functions in the solution of integral equations. Wavelets, which are constructed from a rigorous mathematical foundation, were shown by Steinberg and Leviatan[7], and by Wang[8] to have great promise of success in converting an integral equation into a sparse matrix equation. These authors used wavelets to solve Fredholm integral equations of the first kind for electromagnetic scattering problems.

In this paper, we apply a wavelet expansion to the volume integral equation for eddy current modeling using a Fredholm integral equation of the second kind derived from the diffusion equation. A preliminary study of this problem was presented by Wang et al.[9] Results from the wavelet method were compared with results from a layer approximation and with experimental results by Satveli at al.[10] In this paper, we give a comprehensive discussion of the wavelet expansion method with both theoretical derivations and implementation details. A new linear system of equations is obtained using the wavelet expansion in a Galerkin method to solve the governing volume integral equation. By using a proper threshold, a sparse system matrix can be generated which results in a large savings in execution time and memory requirements. Both wavelets on the real line and periodic wavelets were used. We also provide some implementation details which are important when applying this method to real-world problems.

To evaluate the performance of the wavelet method, we present several numerical examples with the Haar wavelet[11] and the Daubechies' compactly supported wavelet[11] with a periodic extension. Numerical examples show how wavelets change the sparsity of the system matrix. We also observed an interesting relationship between the physical nature of the eddy current problem and the structure of the system matrix. To evaluate the accuracy lost in the process of applying the wavelet transform and thresholding, we conducted an experiment on a simulated corrosion pit underneath a thin aluminum plate. The experimental results are compared with results given by the wavelet method. We observed that a very high

compression rate can be achieved with only a small loss in accuracy, which is often negligible.

The organization of the paper is as follows: In Sec. 2 we briefly review the volume integral method for eddy current modeling. In Sec. 3, we apply wavelet expansions to the eddy current modeling problem and derive a new system of linear equations after using a wavelet basis in the Galerkin method. In Sec. 4, we discuss important implementation details for applying the method, give numerical examples and then compare the theoretical result from the wavelet method with experimental results. Finally, the paper is concluded with a summary in Sec. 5.

2. The Volume Integral Method for Eddy Current Modeling

In this section, we briefly review the volume integral method used in eddy current modeling. The volume integral method was originally used in geophysical induction studies and then introduced to the field of eddy current NDE by Dunbar.[3] This method transforms the governing eddy-current diffusion equation to an integral equation suitable for numerical solution using a Green's function for the given geometry. This volume integral equation can be discretized and transformed to a linear matrix equation for solving directly.

It is well known that in the frequency range of eddy current testing, the displacement current is very small and can be neglected (quasistatic assumption). In this case Maxwell's equations reduce to a diffusion equation. The volume integral method uses the following volume integral equation[3]

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) + \int_{V'} \delta\sigma(\mathbf{r}')\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}')dv' \qquad (1)$$

for solving the diffusion equation where $\mathbf{r}'$ is the source coordinate, $\mathbf{r}$ is the field coordinate, V' is the flaw volume, and $\delta\sigma(\mathbf{r}')$ is the conductivity change between the flaw and the host material. $\mathbf{E^0(r)}$ is the incident electric field, when there is no flaw in the test piece. Analytical solutions exist for the incident field for an air-core coil above a half-space or layered metal.[12] $\mathbf{E(r)}$ is the total electric field in the presence of a flaw. $\mathbf{G(r,r')}$, the Green's function for the given test geometry, is a special solution of the governing diffusion equation

$$\nabla^2\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')-i\omega\mu\sigma_0\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')=i\omega\mu\delta(\mathbf{r}-\mathbf{r}')\mathbf{I}, \tag{2}$$

where σ_0 is the host conductivity, $\delta(\mathbf{r}-\mathbf{r}')$ is the three-dimensional Dirac delta function, and $\mathbf{I}$ is the unit dyad.

Having solved for the total field using (1), the impedance change of the coil next to a non-magnetic metal can be obtained from the reciprocity formula[13]

$$\Delta Z=-\frac{1}{I^2}\int_{V'}\delta\sigma(\mathbf{r}')\mathbf{E}^0(\mathbf{r}')\cdot\mathbf{E}(\mathbf{r}')dv', \tag{3}$$

where I is the current in the coil, and ΔZ is the change in impedance caused by the change in conductivity $\delta\sigma(\mathbf{r}')$. This equation is particularly useful because it only requires the incident field and the total field to be evaluated within the flaw volume. It is not necessary to numerically compute the fields in the region of the coil.

Given the Green's function and the incident field $\mathbf{E}^0$, the solution of the total field $\mathbf{E}$ is usually accomplished by discretizing (1). The flaw volume is divided into N smaller volume elements within which the electric field can be assumed constant. After discretization, (1) becomes

$$\mathbf{E}^0(\mathbf{r}) = \mathbf{E}(\mathbf{r}) + \sum_{i=1}^{N} \int_{V_i} \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \delta\sigma_i dv_i \cdot \mathbf{E}_i, \tag{4}$$

which can be further reduced to the following linear system

$$\mathbf{E}^0 = (\mathbf{I} - \mathbf{C})\mathbf{E}, \tag{5}$$

where $\mathbf{E}^0$ is incident field, $\mathbf{E}$ is the total field, $\mathbf{I}$ is the identity matrix and $\mathbf{C}$ is a matrix derived from the conductivity change and the integration of the Green's function in each element.

The solution of the above linear system can be obtained by using LU decomposition and back substitution. However, computational difficulties arise when the number of elements N is large. It is easy to see that the memory requirement for storing the system matrix is proportional to N^2, and the number of operations required to invert the system matrix is on the order of N^3. Also, the accuracy of the method is related to the sizes of the volume elements. Because in metal the electric field change is closely related to the skin depth, it is necessary for the dimension of each volume element be much smaller than the skin depth to accurately model the field distribution. For a large (in units of skin depth) three-dimensional flaw, the number of elements required will be on the order of thousands. Such a large number of element causes a very high, sometimes almost unattainable, computational requirement for the computing resource. Therefore, to apply the volume integral method to large flaws, we must find a way to reduce the space and time requirements. It is in this sense that wavelet analysis provides an ideal tool for us to reduce the redundancy in the computation and compress the size of the computational burden.

3. Wavelet Expansions in the Solution of Volume Integral Equations

In this section we apply a wavelet basis to the solution of the three-dimensional volume integral equation introduced in Section 2. Before discussing the wavelet expansion method, we briefly introduce orthonormal wavelet basis and periodic wavelet basis. To simplify the mathematical derivation, we consider the one-dimensional case first and derive the new linear system equations by using the Galerkin method. Then we extend the result to the three-dimensional case. The discussion will be based on using a wavelet basis on the real line. We will consider the modification of the method when using periodic wavelets.

3.1 Orthonormal Wavelet Basis and Periodic Wavelet Basis

This discussion of the orthonormal wavelet basis and the periodic wavelet basis is limited to the fundamental definitions and aspects of wavelets that are essential for the completeness of this paper. For detailed discussions on wavelet theory, refer to the books by Daubechies,[11] Chui[14] and Meyer,[15] and the papers by Strang[16] and Mallat.[17]

Given a mother wavelet $\psi(x)$, the wavelets

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \tag{6}$$

form an orthonormal basis if

$$\langle \psi_{j,k}, \psi_{l,m} \rangle = \delta_{j,l}\delta_{k,m}, \text{ for all } j,k,l,m \in Z \tag{7}$$

where $\langle f, g \rangle$ is the inner product operation of functions $f(x)$ and $g(x)$ in $L^2(R)$, and is defined by

$$\langle f, g\rangle = \int_{-\infty}^{+\infty} f(x)g(x)dx, \tag{8}$$

Z is the set of integers and $\delta_{p,q}$ is the Kronecker delta function. The simplest orthonormal wavelet, the Haar wavelet, and its corresponding scaling function are shown in Fig. 1. Another orthonormal wavelet, the Daubechies compactly supported wavelet with four coefficients, and its corresponding scaling function are shown in Fig. 2.

Given any function $f(x)$ in $L^2(R)$, we can approximate it by using an orthonormal wavelet basis

$$f(x) = c + \sum_j \sum_k a_{j,k}\psi_{j,k}(x), \tag{9}$$

where c is a constant, and the expansion coefficients $a_{j,k}$ can be obtained by projecting $f(x)$ onto the subspace expanded by $\psi_{j,k}(x)$ as

$$a_{j,k} = \langle f(x), \psi_{j,k}(x)\rangle. \tag{10}$$

Given an orthonormal wavelet $\psi(x)$ on the real line, the periodic, orthonormal wavelets in $[0, 1]$ can be defined as[11]

$$\psi_{j,k}^{per}(x) = \sum_{n\in Z} \psi_{j,k}(x+n) \tag{11}$$

where $j, k, n \in Z$. It can be shown that $\{1\}$ and $\{\psi_{j,k}^{per} \mid j = 0, 1, 2, \cdots, k = 0, \cdots, 2^j - 1\}$ constitute a periodic, orthonormal basis in $[0, 1]$.[11]

If the orthonormal wavelet in equation (6) is compactly supported, the infinite summation in equation (11) reduces to finite summations. The periodic, orthogonal wavelet $\psi_{0,0}^{per}(x)$ constructed from the Daubechies compactly supported wavelet with four coefficients is shown in Fig. 3.

3.2 One Dimensional Case

Consider the following one dimensional simplification of the problem defined in equation (1)

$$E^0(x) = E(x) - \int_\Omega G(x,x')\delta\sigma(x')E(x')dx', \tag{12a}$$

$$E(x) = 0, \qquad x \notin \Omega, \tag{12b}$$

where $\Omega = [0, L]$ is the solution region which contains the flaw. The explicit boundary condition of equation (12b) is the result of using the wavelets on the real line, since some of the wavelet basis functions used for the expansion of the unknown total field $E(x)$ in the solution region Ω do not vanish outside Ω. Now we use the Galerkin method to approximate $E(x)$ in Ω with an orthonormal wavelet basis on the real line giving

$$E(x) = \frac{a_0}{\sqrt{L}} + \sum_{j=j_1}^{j_2} \sum_{k=k_1(j)}^{k_2(j)} a_{jk}\psi_{j,k}(x), \qquad x \in \Omega, \tag{13}$$

where a_0 and a_{jk} are the expansion coefficients and $\sqrt{L}$ is a normalization factor. The expansion of $E(x)$ uses wavelets from resolution level j_1, which corresponds to the lowest resolution component of $E(x)$, to the resolution level j_2, which corresponds to the desired resolution for the solution. Notice that equation (13) is only satisfied when $x \in \Omega$. Setting

$$p(x)=\begin{cases}1 & x\in\Omega\\ 0 & x\notin\Omega,\end{cases} \tag{14}$$

we can express the total field on the real line as

$$E(x)=\frac{a_0 p(x)}{\sqrt{L}}+\sum_{j=j_1}^{j_2}\sum_{k=k_1(j)}^{k_2(j)} a_{jk}\psi_{j,k}(x)p(x). \tag{15}$$

To simplify the expression, we can combine the scaling index and translation index of the basis functions into one index. Let

$$\psi_0(x)=\frac{p(x)}{\sqrt{L}} \tag{16}$$

and suppose the number of basis functions used in the summation of (15) is $N-1$, the total field can be expressed as

$$E(x)=\sum_{i=0}^{N-1} a_i\psi_i(x)p(x), \tag{17}$$

where $\{a_i\}$, $\{\psi_i(x)\}$ are derived from $\{a_{jk}\}$, $\{\psi_{j,k}(x)\}$ by combining the indices. Substituting equation (17) into equation (12a) gives

$$E^0(x)=\sum_{i=0}^{N-1} a_i\psi_i(x)p(x)+\int_\Omega G(x,x')\delta\sigma(x')\sum_{i=0}^{N-1} a_i\psi_i(x')p(x')dx'. \tag{18}$$

Since $\delta\sigma(x')=0$ if $x\notin\Omega$, we can rewrite the right hand side of equation (18) as

$$E^0(x) = \sum_{i=0}^{N-1} a_i \psi_i(x) p(x) + \sum_{i=0}^{N-1} a_i \langle G(x,x')\delta\sigma(x'), \psi_i(x') \rangle. \tag{19}$$

Taking the inner product of both sides of equation (19) with $\psi_j(x), j = 0, 1, \cdots, N-1$ gives

$$\langle E^0(x), \psi_j(x) \rangle = \sum_{i=0}^{N-1} a_i \langle \psi_i(x) p(x), \psi_j(x) \rangle + \sum_{i=0}^{N-1} a_i \langle \langle G(x,x')\delta\sigma(x'), \psi_i(x') \rangle, \psi_j(x) \rangle, \tag{20}$$
$$j = 0, 1, \cdots, N-1.$$

These linear equations can be written in more compact form as

$$\sum_{i=0}^{N-1} a_i (D_{ij} + C_{ij}) = b_j, \quad j = 0, 1, \cdots, N-1 \tag{21}$$

where $b_j = \langle E^0(x), \psi_j(x) \rangle$ is the wavelet expansion coefficient of the incident field, $C_{ij} = \langle \langle G(x,x')\delta\sigma(x'), \psi_i(x') \rangle, \psi_j(x) \rangle$ is the two-dimensional wavelet expansion coefficient of the function $G(x,x')\delta\sigma(x')$, and $D_{ij} = \int_{\Omega} \psi_i(x)\psi_j(x)dx$. Given the Green's function for the test geometry, the incident field distribution and the conductivity distribution in the flaw region, b_j and C_{ij} can be easily obtained by using the fast wavelet transform algorithm, and D_{ij} is derived directly from the wavelet basis selected.

Equation (21) is the general form of the linear system equations for the solution of equation (12a) after using an orthonormal wavelet basis on the real line. The reason why D_{ij} does not reduce to the Kronecker delta function is that the basis functions $\{\psi_i(x)\}$ used to approximate the total field are not always orthogonal to each other in Ω. Suppose the support of $\psi_i(x)$ is S_i, and the support of $\psi_j(x)$ is S_j, then we have

$$D_{ij} = \delta_{ij} \qquad \text{if} \quad S_i \cap S_j \subseteq \Omega. \tag{22}$$

Therefore, if any part of the overlapped region of S_i and S_j are outside of Ω, the value of D_{ij} must be calculated explicitly. This is another disadvantage of using wavelets on the real line in the solution of equation (12a) in a bounded region. A special case of (21) exists when we use the Haar wavelet to approximate the unknown total field. In this case, we are able to express the total field in Ω by using basis functions all lying in Ω because the Haar wavelet has a support as small as the grid separation, and $D_{ij} = \delta_{ij}$ because $S_i \subseteq \Omega., i = 0, 1, \cdots, N-1$. Therefore, in the case of the Haar basis, there is no problem at the boundary of the solution region and the boundary condition of equation (12b) is not necessary. However, since the Haar basis is not efficient in approximating smooth functions, the resultant system matrix may not be as sparse as that from a smoother wavelet.

3.3 Extension to the Three-Dimensional Case

The above result can be extended to the three-dimensional case by a similar, but somewhat lengthy derivation. The three-dimensional volume integral problem is given by

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) + \int_{V'} \delta\sigma(\mathbf{r}')\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}')dv', \tag{23a}$$

$$\mathbf{E}(\mathbf{r}) = \mathbf{0}, \qquad \mathbf{r} \notin V'. \tag{23b}$$

We can construct an orthonormal basis for $L^2(R^3)$ by using the tensor product functions generated by three one-dimensional orthonormal wavelet bases

$$\psi_{j_x,k_x;j_y,k_y;j_z,k_z}(x,y,z) = \psi_{j_x,k_x}(x)\psi_{j_y,k_y}(y)\psi_{j_z,k_z}(z), \tag{24}$$

where the wavelets in x, y and z directions are scaled independently.

By a similar formulation and combination of indices as what we used in the section above, we can express the total vector field as

$$\mathbf{E}(\mathbf{r}) = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \sum_{k=0}^{N_z-1} \mathbf{a}_{ijk} \psi_{i;j;k}(x,y,z) p_x(x) p_y(y) p_z(z) \tag{25}$$

where $\mathbf{a}_{ijk}$ is the vector expansion coefficient for the three-dimensional total vector field $\mathbf{E}(\mathbf{r})$.

Substituting equation (25) into equation (23a) and using the same procedure as in the section above, we can obtain the linear system equations for the three-dimensional case as

$$\sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \sum_{k=0}^{N_z-1} \mathbf{a}_{ijk} \cdot (\overline{\mathbf{D}}_{ijklmn} + \overline{\mathbf{C}}_{ijklmn}) = \mathbf{b}_{lmn} \tag{26}$$

$$l = 0, 1, \cdots, N_x - 1, \quad m = 0, 1, \cdots, N_y - 1, \quad n = 0, 1, \cdots, N_z - 1,$$

where $\mathbf{b}_{lmn}$ is the vector expansion coefficient for the incident vector field, $\overline{\mathbf{C}}_{ijklmn}$ is the 3 by 3 matrix expansion coefficient of the six-dimensional function $\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\delta\sigma(\mathbf{r}')$ and $\overline{\mathbf{D}}_{ijklmn}$ is a 3 by 3 matrix with 9 identical elements, and each element can be expressed as

$$d_{ijklmn} = \int_{V'} \psi_{i;j;k}(x,y,z) \psi_{l;m;n}(x,y,z) dv' \tag{27}$$

3.4 Extension to the Periodic Wavelet Case

As we have seen in Section 3.2, using wavelets on the real line in the solution of integral equations like equation (12a) has two major disadvantages. First, the boundary condition of equation (12b) must be explicitly enforced. This is a very undesirable requirement in applying this method since the total field generally does not vanish outside the flaw region. To satisfy equation (12b), we must enlarge the solution region Ω so that outside Ω the total field is very small and can be neglected. This may result in a solution region which is much larger than the flaw region and thus reduce the usefulness of this method. Another major disadvantage is that because some of the wavelet basis functions used to approximate the total field are not orthogonal in the solution region, the term D_{ij} does not reduce to delta functions and must be calculated before the solution of equation (21). These disadvantages all arise because the fact that wavelet basis functions on the real line used to approximate the total field do not form an orthogonal set in the solution region Ω. The periodic wavelet introduced in Section 3.1, on the other hand, can be used to avoid these difficulties because it can be made orthonormal in a bounded region Ω. Using an orthonormal, periodic wavelet basis $\left\{\psi_{j,k}^{per}(x)\right\}$ defined in $\Omega = [0, L]$, we can expand the total field in Ω as

$$E(x) = \sum_{i=0}^{N-1} a_i \psi_i^{per}(x), \qquad x \in \Omega \tag{28}$$

by combining indices. Notice that the right hand side of equation (28) denotes a periodic function with period equal to L and whose value is equal to the total field in region Ω. Substituting equation (28) into equation (12a), we have

$$E^0(x) = \sum_{i=0}^{N-1} a_i \psi_i^{per}(x) + \int_{\Omega} G(x,x')\delta\sigma(x') \sum_{i=0}^{N-1} a_i \psi_i^{per}(x')dx', \qquad x \in \Omega \tag{29}$$

Projecting both sides of equation (30) into the subspaces expanded by $\psi_j^{per}(x),\ j=0,1,\cdots,N-1,$ we have the following equations

$$\int_{\Omega} E^0(x)\psi_j^{per}(x)dx = \sum_{i=0}^{N-1} a_i\delta_{ij} + \sum_{i=0}^{N-1} a_i \iint_{\Omega\Omega} G(x,x')\delta\sigma(x')\psi_i^{per}(x')\psi_j^{per}(x)dx'dx, \qquad (30)$$
$$j=0,1,\cdots,N-1,$$

which can be rewritten in the form of a matrix equation

$$\sum_{i=0}^{N-1} a_i(\delta_{ij}+C_{ij}) = b_j, \quad j=0,1,\cdots,N-1 \qquad (31)$$

where b_j are the periodic wavelet expansion coefficients of the incident field and C_{ij} are the periodic wavelet expansion coefficients of $G(x,x')\delta\sigma(x')$. The use of periodic wavelets in the solution of equation (12a) gives a group of well formed linear equations: all coefficients in equation (31) can be obtained by using the fast wavelet transform algorithm of Mallat.[17] Also, since the boundary condition of equation (12b) is not necessary, we can use a solution region which has the same size as the flaw region.

Despite its advantage over the wavelet on the real line, a potential disadvantage of using periodic wavelets exists due to the fact that the periodization of the total field may introduce an edge effect. Since the values of the total field at two ends of the solution region generally are not equal, there are discontinuous points at the boundaries of each period. This discontinuity introduces some large expansion coefficients for the total field and makes the system matrix less sparse. However, experience shows that this edge effect generally does not have a large influence on the overall performance of the wavelet method.

4. Numerical Results and Comparison with Experimental Results

To evaluate the validity and performance of the wavelet method discussed in Section 3, we implemented the method and used the problem of hidden corrosion in an aluminum plate as a test case. In Section 4.1, we describe the experimental setup for the model problem used as the test case for the wavelet method and in Section 4.2 we give some implementation considerations of the wavelet method. In Section 4.3, we give the numerical results obtained by using the Haar wavelet and the periodic wavelet constructed from the Daubechies compactly supported wavelets with four coefficients and compare the results from the wavelet method with the experimental result. In Section 4.4, we discuss some relationships between the physical nature of the eddy current problem and the selection of wavelet basis in the solution of the volume integral equation.

4.1 The Experiment

The test case is to determine the change in the impedance of a right-cylindrical air-core coil next to an aluminum plate that contains a right-cylindrical flat-bottom hole on the side opposite the coil. The half space below the coil can be considered as a two layer structure, where the first layer is the aluminum plate and the second layer is vacuum. A schematic drawing of the test is shown in Fig. 4. In the experiment, the impedance change of a precision wound coil of copper wire was measured at 20 frequencies, equally-spaced between 2.5 kHz and 50 kHz, with an HP 4194A impedance analyzer. The position of the coil was held fixed, while the position of the sample was varied by a precision x-y scanning table that translated the sample perpendicular to the axis of the coil. Scans were repeated five times and the results averaged to produce the results reported herein. Measurements of the coil impedance were made as a function of frequency and position with respect to the center of the flat-bottom hole.

The coil has a right-cylindrical cross-section and consisted of 504 turns of copper

wire wound on an insulating nonmagnetic core. The inner radius of the coil is 3.8 mm, the outer radius is 5.63 mm, and the height of the coil is 2.42 mm. The distance between the bottom of the coil and the top of the plate is called "lift-off" and is 0.23 mm.

The test sample consists a thin flat plate of 2024 aluminum alloy. The dimensions of the plate are 75 mm x 125 mm x 1.02 mm. A nominally right-cylindrical hole was drilled on the side of the plate opposite the coil. The diameter of the hole was 6.30 mm and the depth ranged between 0.28 mm at the center to 0.27 mm at the edge. This flaw is relatively large compared with the skin depth in the test frequency range and direct modeling by the volume integral method is not feasible due to the large computational resource requirement.

4.2 Implementation of the Wavelet Method

The modeling problem described in Section 4.1 can be considered as the problem of determining the impedance change of the coil caused by a flaw in an infinite flat metal plate because the aluminum plate is very thin and the flaw is not close to the edges of the plate. The analytical solution for the incident field of an air-core coil in an infinite thin plate has been given by Dodd and Deeds.[12] The Green's function for the thin plate geometry has also been derived by Weaver.[18] The fast wavelet transform algorithm used in our implementation is Mallat's pyramid algorithm for an orthonormal wavelet basis.[17] It is worth mentioning that all the electric field quantities in the implementation are vectors with complex values. The wavelet transform of a complex function is obtained by transforming the real and imaginary parts independently and then combining the results. Also, to obtain $\mathbf{b}_{lmn}$ and $\overline{\mathbf{C}}_{ijklmn}$ in equation (26), multi-dimensional wavelet transforms must be used. Based on the approach we used to define the three-dimensional wavelets in equation (24), the multi-dimensional wavelet transform can be obtained by transforming the data in each dimension sequentially. The result is independent of the order of the dimensions used in the transforms. This is exactly the same as for the multi-dimensional Fourier transform.

There exists another problem with the storage of the system matrix in the implementation of the wavelet method. The system matrix is the six-dimensional wavelet transform of the product of the Green's function and the conductivity distribution. The problem is how to compute the wavelet transform of the system matrix without fully storing the matrix in the computer memory since it could be very large when the number of volume elements used in modeling is sufficient for an accurate solution. To solve this problem, we utilized the property of Mallat's pyramid algorithm for the fast wavelet transform[17] that the computation at any resolution level is solely based on the result of the immediately higher resolution level. Thus, we are able to compute a wavelet transform of size N from the results of wavelet transforms with a smaller size M if both N and M are powers of 2. Based on this relationship, we can compute the wavelet transform of the whole system matrix in two steps. In the first step, we divide the system matrix into many submatrices with identical sizes and compute the wavelet transform for each of them. The result is stored in a sparse matrix data structure. In the second step, the wavelet transform of the whole matrix is obtained by using the result of the first step. The result is then thresholded and stored back in the sparse matrix data structure.

The steps used to implement the wavelet method are summarized as follows:

1. Compute the incident field distribution by using Dodd and Deed's result.
2. Compute the integral of the Green's function in each volume element. The Green's function for thin plate geometry is calculated by using the result of Weaver. For singular elements, both analytical and numerical integrations are used to stabilize the result. For nonsingular elements, only numerical integration is used.
3. Compute the wavelet transform of the system matrix by using the two-step algorithm mentioned above.
4. Invert the transformed system matrix by using a sparse matrix LU decomposition routine.
5. Compute the wavelet transform of the incident field.
6. Compute the wavelet coefficients of the total field by backsubstitution.
7. Take the inverse wavelet transform to obtain the total field.

8. Calculate the impedance change by using the reciprocity principle formula (equation 7).

4.3 Comparison of the Numerical Result and the Experimental Result

Based on the procedure introduced in Section 4.2, two numerical tests were performed for the test case. In the first numerical test, the aim was to compare the result obtained from the wavelet method by using the Haar basis and the periodic wavelet basis constructed from the Daubechies compactly supported wavelet with four coefficients, which we will refer to as DAUB4P in later discussions for simplicity. In the first test, the test frequency was selected as 10 kHz, which corresponds to a skin depth of 1.17 mm in 2024 aluminum. The solution region was a rectangle with a length and a width equal to the diameter of the cylindrical flaw and a height equal to the height of the flaw. The solution region was modeled by 512 rectangular elements with 16 elements along each of the two horizontal directions and 2 elements in the vertical direction. In the first comparison, the system matrices were obtained by using a threshold of 0.005, which is relatively small, to ensure that a good result would be obtained. The matrix map for the sparse matrix obtained by using the Haar basis is shown in Fig. 5. After thresholding, only 4.36% of the total number of the matrix elements were nonzero elements. The matrix map for the sparse matrix obtained by using DAUB4P is shown in Fig. 6. In this case, 5.05% of all the matrix elements were nonzero elements after thresholding. The impedance change results obtained by using the Haar basis and DAUB4P are compared with the experimental results in Fig. 7. The theoretical results are in good quantitative agreement with the experimental results. Also there is no significant difference between the result obtained using the Haar wavelet and the result obtained using the periodic wavelet in this case.

To examine the effect of the threshold on the accuracy of the results, or in other words, to decide whether there is some range of the threshold within which the results remain almost unchanged, we recomputed the results for the numerical example above using a threshold of 0.05 for zeroing the matrix elements. The matrix map for the sparse matrix obtained by using

the Haar basis is shown in Fig. 8. After thresholding, only 0.413% of the matrix elements were nonzero. The matrix map for the sparse matrix obtained by using the DAUB4P is shown in Fig. 9. In Fig. 9, only 0.254% of the matrix elements were nonzero after thresholding. One can see that after using a threshold of 0.05, most nonzero elements are on the diagonal of the system matrix. The impedance change results in this case are compared with the experimental results in Fig. 10. Although the threshold is ten times larger and the thresholded matrices are rather sparse, an overall correspondence between the theoretical results and the experimental result can still be observed. It can be seen from this example that the wavelet method does indeed have a good potential for reducing the redundancy of the problem.

The second numerical test was to examine the stability of the method over a relatively large frequency range. The theoretical impedance changes using DAUB4P and a threshold of 0.005 were computed for the frequency range from 2.5 kHz to 50 kHz, with a stepsize of 2.5 kHz, and for coil-flaw offsets from 0 mm to 8 mm, with a stepsize of 1 mm. The impedance change results as a function of frequency for offsets 0 mm, 4 mm and 8 mm are compared with the experimental results in Fig. 11. It is observed that the agreement between the results from the wavelet method and the experimental results are stable over this relatively large frequency range.

4.4 Selection of Wavelet Basis

The matrix structures shown in Fig. 5 and Fig. 6 indicate some interesting relationships between the physical nature of the eddy current problem and the properties of the wavelet method. First, because the eddy current problem is governed by a diffusion equation, the diffusive incident field and total field are generally smooth over the solution region. This means they have larger low-resolution components than high-resolution components. The use of the wavelet basis can separate the more important low-resolution components from the less important high-resolution components, and thus achieve high compression with small

loss in accuracy. This property of the wavelet method is clearly shown in Fig. 5 and Fig. 6, as there are more nonzero elements located in the upper-left corner of the matrix where low resolution wavelets are located, and fewer located in the bottom-right corner where the high resolution wavelets are located. Second, because the Green's function for the eddy current problem generally decreases exponentially with a rate defined by the skin depth at the test frequency, we expect that the interaction between two basis functions with supports far away from each other be small so that it can be neglected without much loss in accuracy. This is clearly verified in Fig. 5 and Fig. 6, as most nonzero elements are close to the diagonal of the matrix, which indicates a large interaction exists only between the basis functions with overlapped or adjacent supports. The above two observations indicate that the unique spatial-frequency localization property of the wavelet basis provides a significant reduction to the complexity of the eddy current problem which can hardly be achieved by using other basis functions.

The above observations can also be useful in guiding the selection of the mother wavelet to achieve optimal compression. The diffusive nature of the unknown total field requires us to use a smooth wavelet so that the total field can be approximated efficiently. On the other hand, the exponential decay of the Green's function requires us to use a wavelet with small support so that the interaction between basis functions far away from each other can be kept small. However, from wavelet theory, smoothness and compact support are contradictory for orthonormal wavelets.[11] In other words, we are unable to build a smooth orthonormal wavelet with compact support. Therefore, in practice, a tradeoff on smoothness and compact support must be made based on the size of the problem and the model parameters.

5. Summary

We have applied wavelet theory in the solution of the volume integral equation of the eddy current modeling problem, which is a Fredholm integral equation of the second-kind.

The new linear system equations after using the wavelet basis are obtained by applying a Galerkin method to the governing volume integral equation. The derivation is based on the result of the one-dimensional case, followed by the extension to the three-dimensional case and the extension to periodic wavelets. The wavelet method was tested using the problem of modeling hidden corrosion in a thin aluminum plate. Results obtained from both the Haar wavelet and the periodic wavelet constructed from the Daubechies compactly supported wavelet were compared with experimental results. The numerical examples show that by using the wavelet expansion method, it is possible to greatly reduce the computational complexity of the solution of the volume integral equation with about 5-10% loss in accuracy. This makes the eddy-current modeling problem of large three-dimensional flaws solvable using modest computing resources.

Although the wavelet expansion method discussed in this paper has been developed in the context of eddy current nondestructive evaluation, it is directly applicable to other problems which require the solution of a Fredholm integral equation of the second-kind, such as electromagnetic prospecting of the subsurface structure of the earth in geophysics.

6. Acknowledgment

This work is supported by the Center for NDE at Iowa State University. The authors would like to thank Dr. James H. Rose of the Center for NDE for helpful comments on the draft of the paper.

REFERENCES

1. P. C. French and L. J. Bond, J. Nondest. Eval. **7**, 55 (1988).
2. N. Nakagawa and J. Chao, "Extended Magnetic Potential Method for Quasistatic Electromagnetism and Eddy Current Phenomena", in *Review of Progress in Quantitive Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1996), **15**, 339.
3. W. S. Dunbar, J. Nondest. Eval. **5**, 9 (1985).
4. W. S. Dunbar, J. Nondest. Eval. **7**, 43 (1988).
5. D. M. McKirdy, J. Nondest. Eval. **8**, 45 (1989).
6. G. L. Hower and D. E. Hadlock, J. Nondest. Eval. **8**, 247 (1989).
7. B. Z. Steinberg and Y.Leviatan, IEEE Trans. on Antennas Propagat. **41**, 610 (1993).
8. G. Wang, IEEE Trans. on Antennas Propagat. **43**, 170 (1995).
9. B. Wang, J. P. Basart, and J. C. Moulder, "Wavelet expansion in volume integral method of eddy-current modeling," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1996), **15**, 377.
10. R. Satveli, J. C. Moulder, B. Wang, and J. H. Rose, J. Appl. Phys. **79**, 2811 (1996).
11. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
12. C. V. Dodd and W. E. Deeds, J. Appl. Phys. **39**, 2829 (1968).
13. B. A. Auld, "Theoretical characterization and comparison of resonant probe microwave eddy-current methods, " in *Eddy Current Characterization of Materials and Structures*, edited by G. Birnbaum and G. Free (ASTM, Philadelphia, 1981), pp. 332-347.
14. C. K. Chui, *An Introduction to Wavelets*, Academic Press, New York, 1992.
15. Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.
16. G. Strang, SIAM Review **31**, 614 (1989).
17. S. G. Mallat, IEEE Trans. on Pattern Analysis and Machine Intelligence **11**, 674 (1989).
18. J. T. Weaver, Canadian J. of Phys. **45**, 1981 (1967).

Figure Captions

Fig. 1. Haar wavelet and its scaling function, the box function.

Fig. 2. The Daubechies compactly supported wavelet with four coefficients and its scaling function.

Fig. 3. The periodic wavelet $\psi_{0,0}^{per}(x)$ constructed from the Daubechies compactly supported wavelet with four coefficients.

Fig. 4. Schematic drawing of the eddy current test problem of a flat bottom hole on the bottom side of a thin aluminum plate.

Fig. 5. The system matrix map after transformation by using the Haar wavelet and thresholded using a value of 0.005. Black spots denote nonzero elements.

Fig. 6. The system matrix map after transformation by using DAUB4P and thresholded using a value of 0.005. Black spots denote nonzero elements.

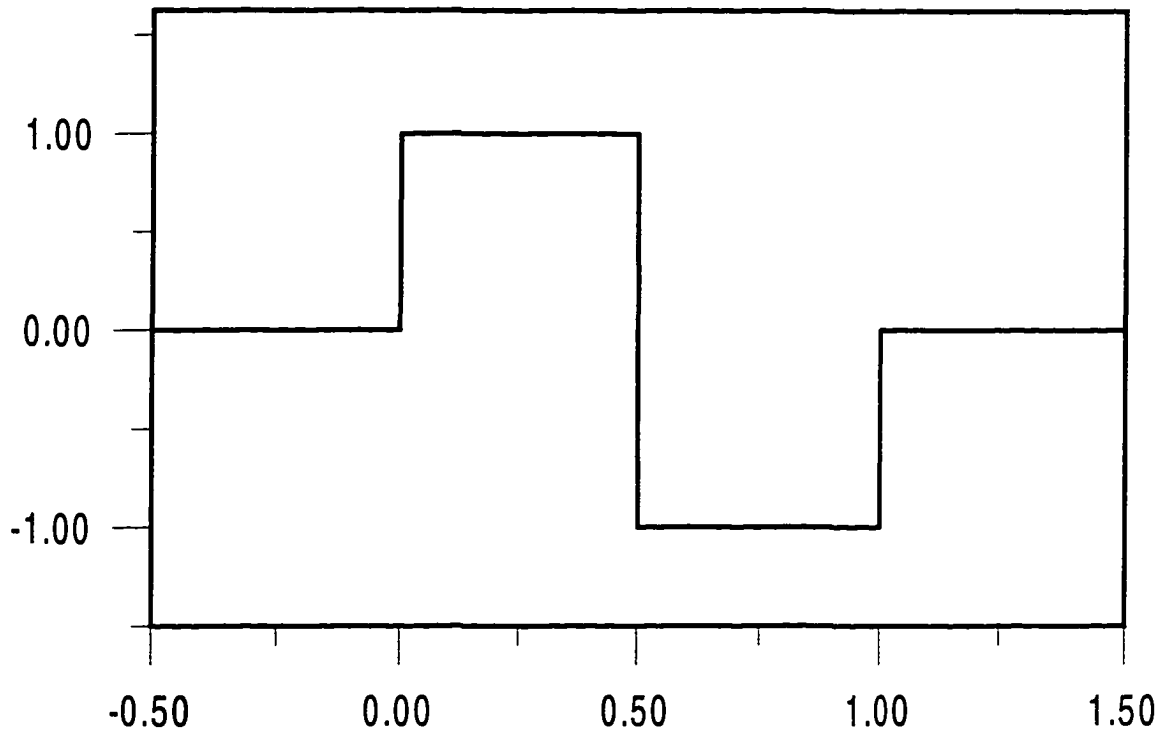
Fig. 7. Comparison of experimental results and the results obtained by using the Haar wavelet and DAUB4P with a threshold of 0.005.

Fig. 8. The system matrix map after transformation by using the Haar wavelet and thresholded using a value of 0.05. Black spots denote nonzero elements.

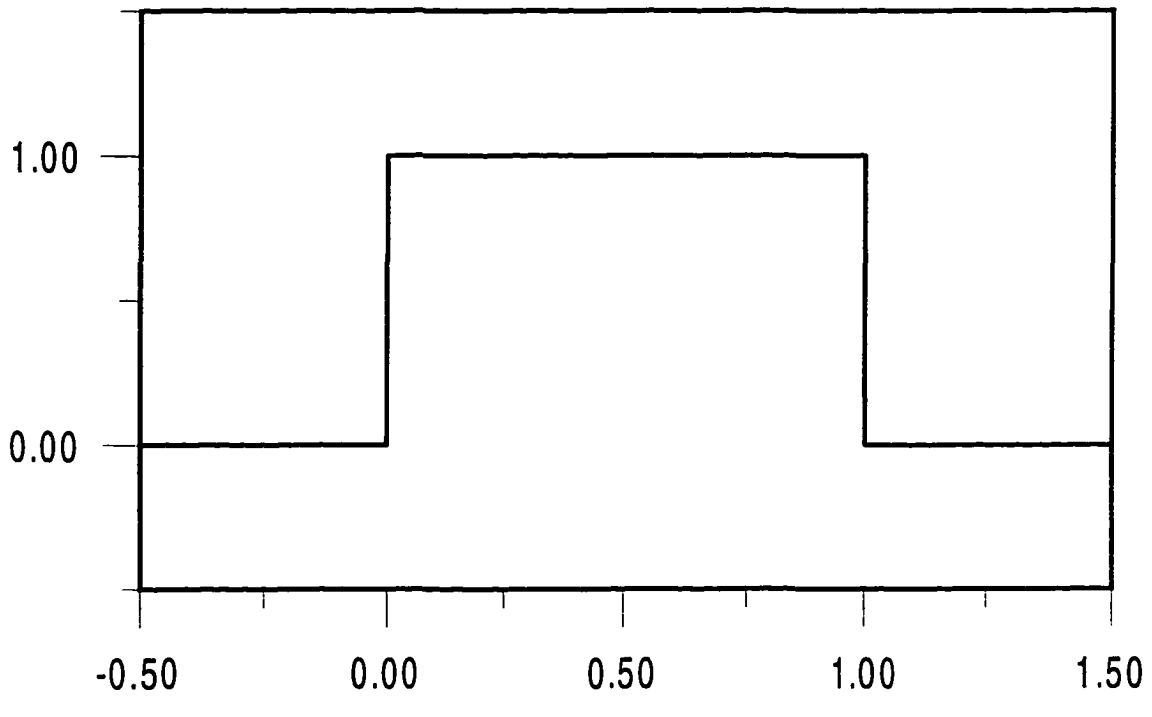
Fig. 9. The system matrix map after transformation by using DAUB4P and thresholded using a value of 0.05. Black spots denote nonzero elements.

Fig 10. Comparison of experimental results and the results obtained by using the Haar wavelet and DAUB4P with a threshold of 0.05.

Fig. 11. Comparison of experimental results and the results obtained by using DAUB4P and a threshold of 0.005 for the frequency range from 2.5 kHz to 50 kHz.

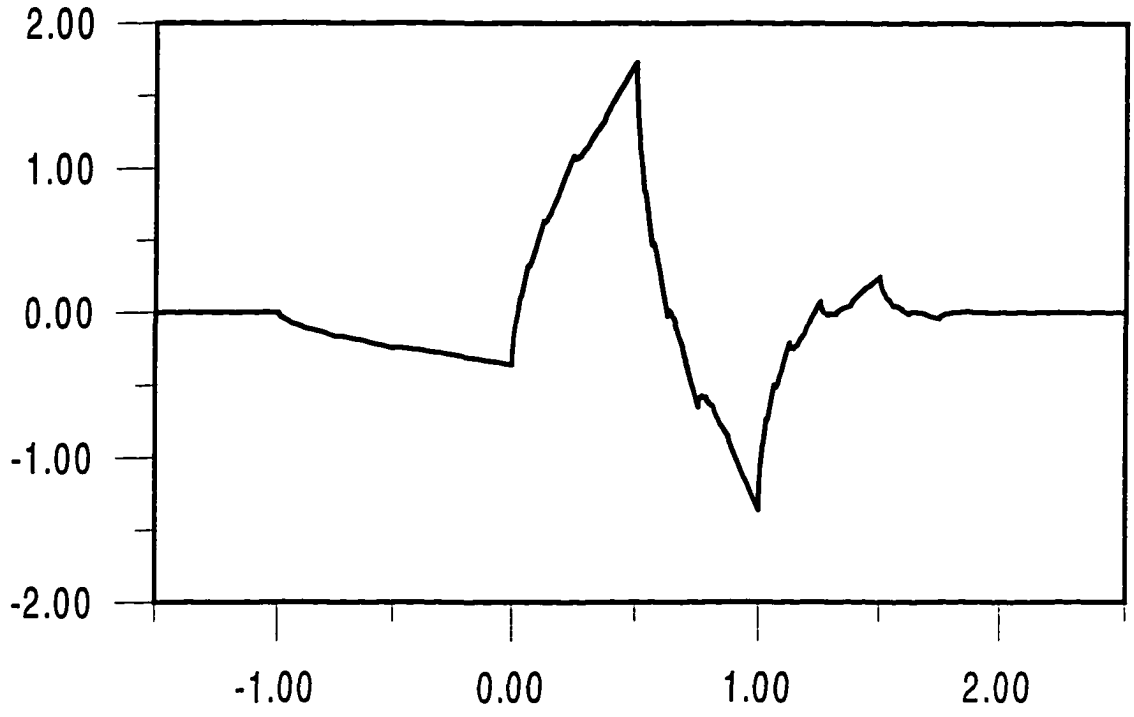


a. Haar wavelet

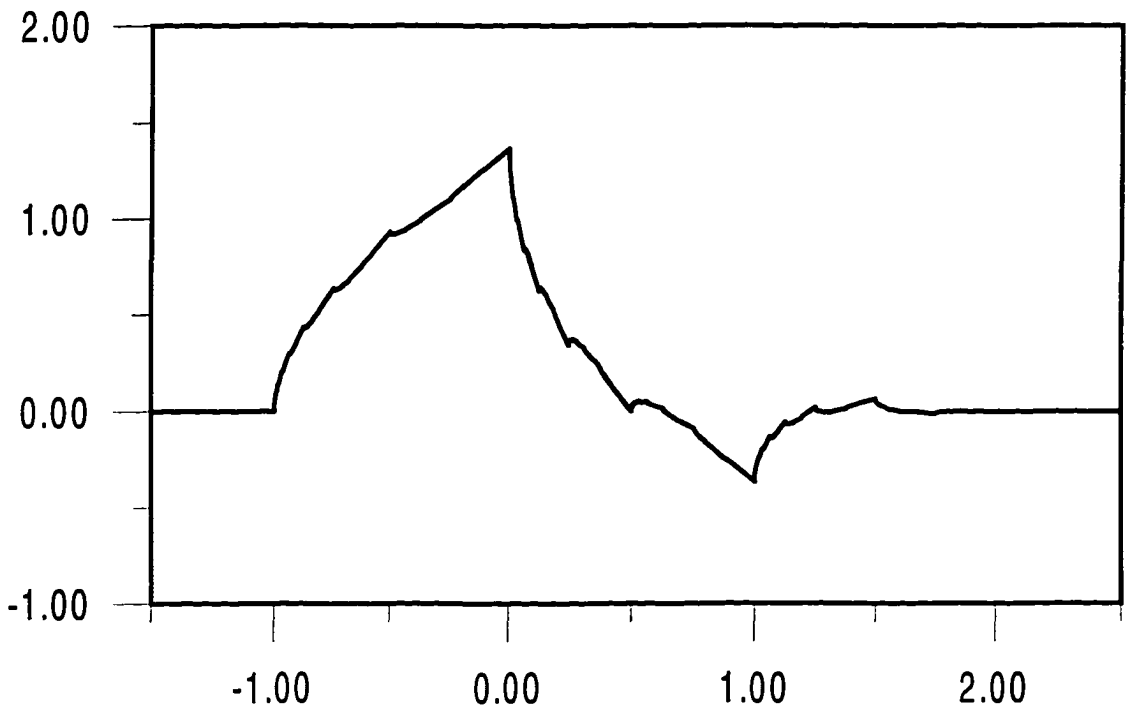


b. The box function

Fig. 1. Haar wavelet and its scaling function, the box function.



a. The Daubechies compactly supported wavelet with four coefficients



b. The corresponding scaling function

Fig. 2. The Daubechies compactly supported wavelet with four coefficients and its scaling function.

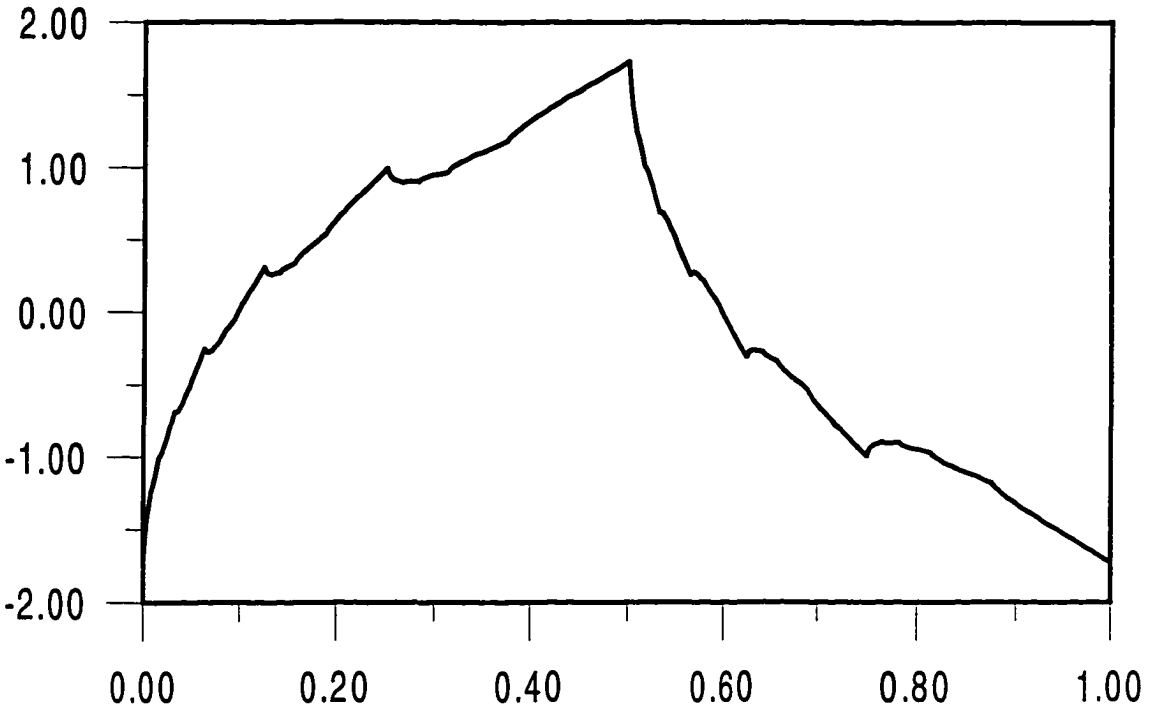


Fig. 3. The periodic wavelet basis $\psi_{0,0}^{per}(x)$ constructed from the Daubechies compactly supported wavelet with four coefficients.

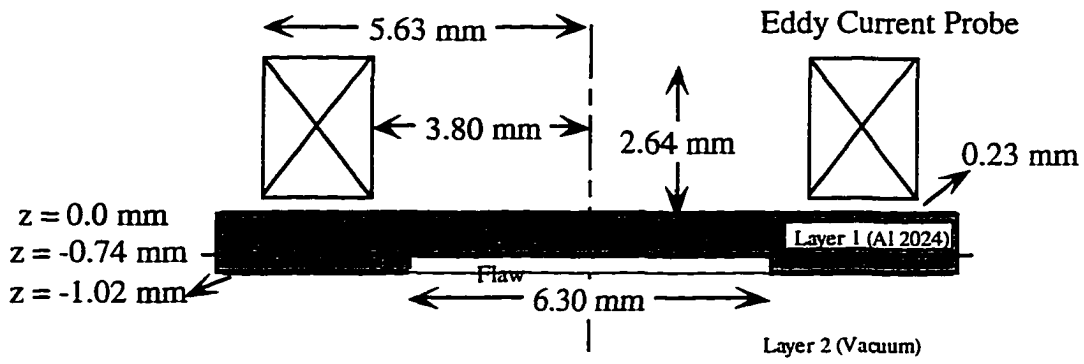


Fig. 4. Schematic drawing of the eddy current test problem of a flat bottom hole on the bottom side of a thin aluminum plate.

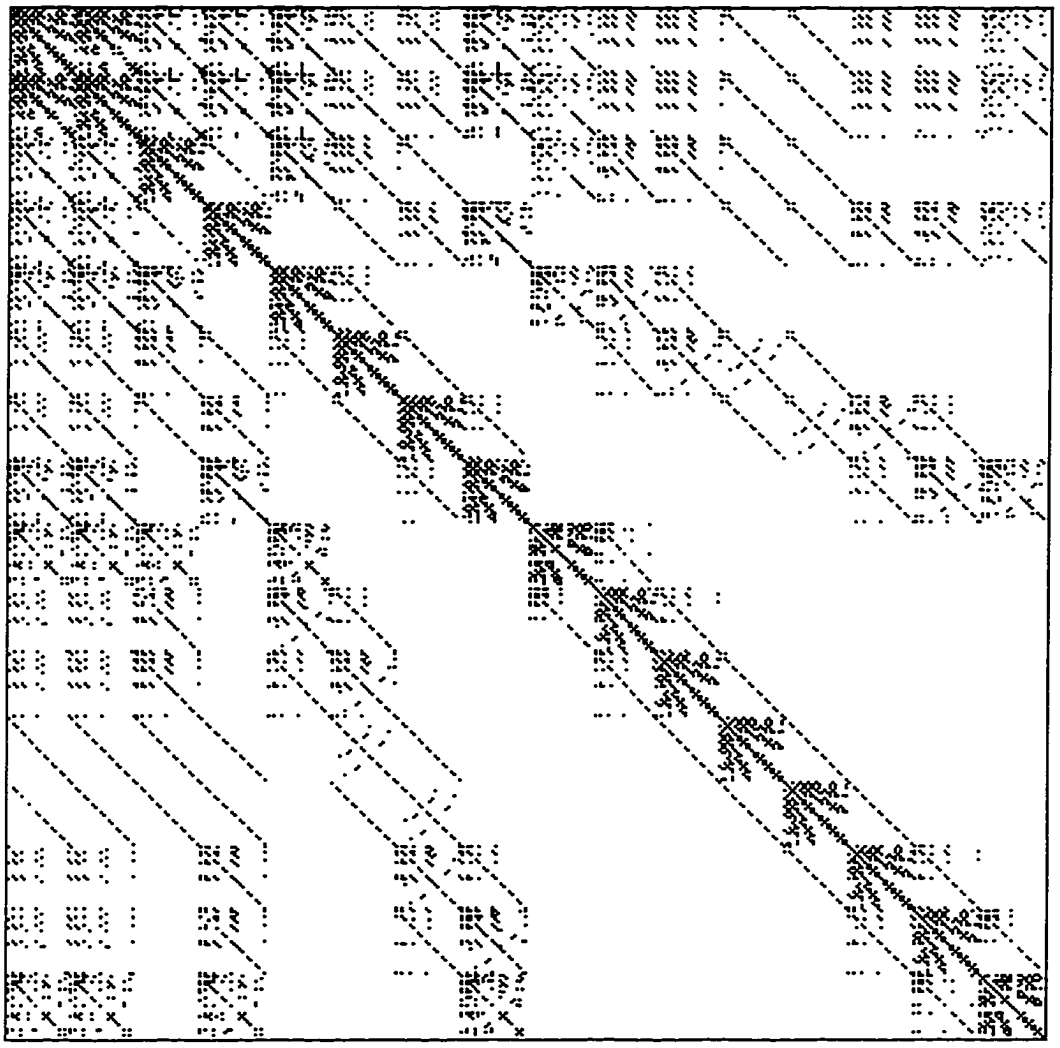


Fig. 5. The system matrix map after transformation by using the Haar wavelet and thresholded using a value of 0.005. Black spots denote nonzero elements.

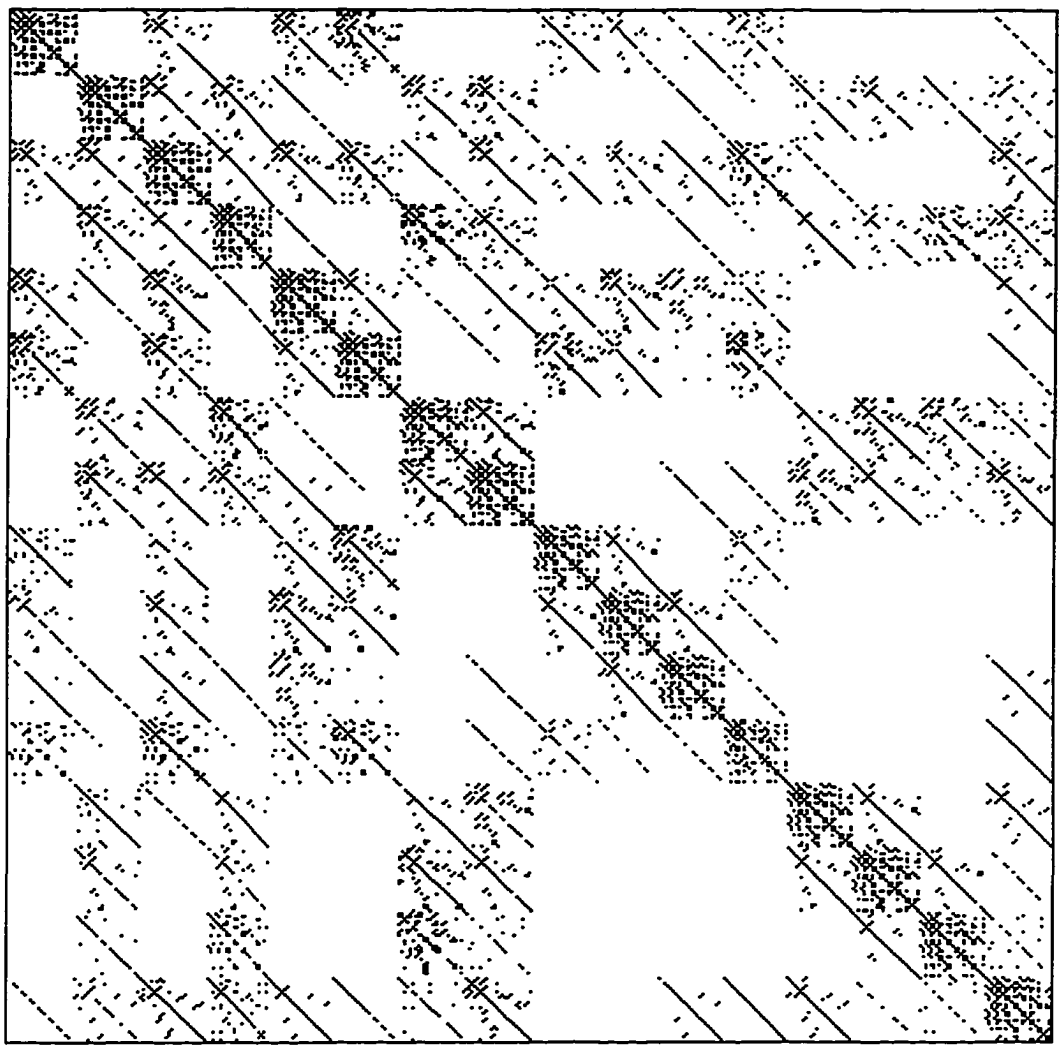


Fig. 6. The system matrix map after transformation by using DAUB4P and thresholded using a value of 0.005. Black spot denotes nonzero element.

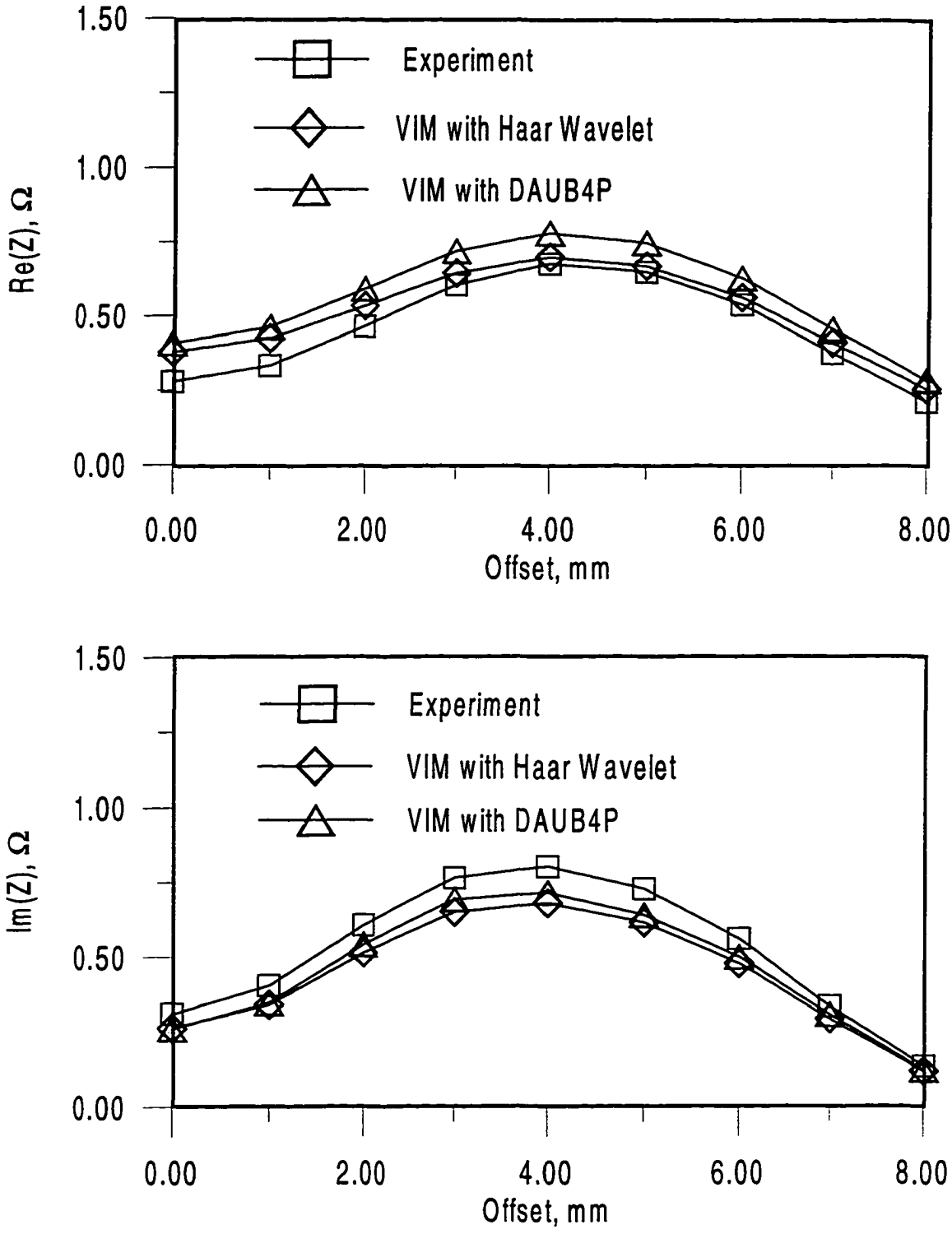


Fig. 7. Comparison of experimental results and the results obtained by using the Haar wavelet and DAUB4P with a threshold of 0.005.

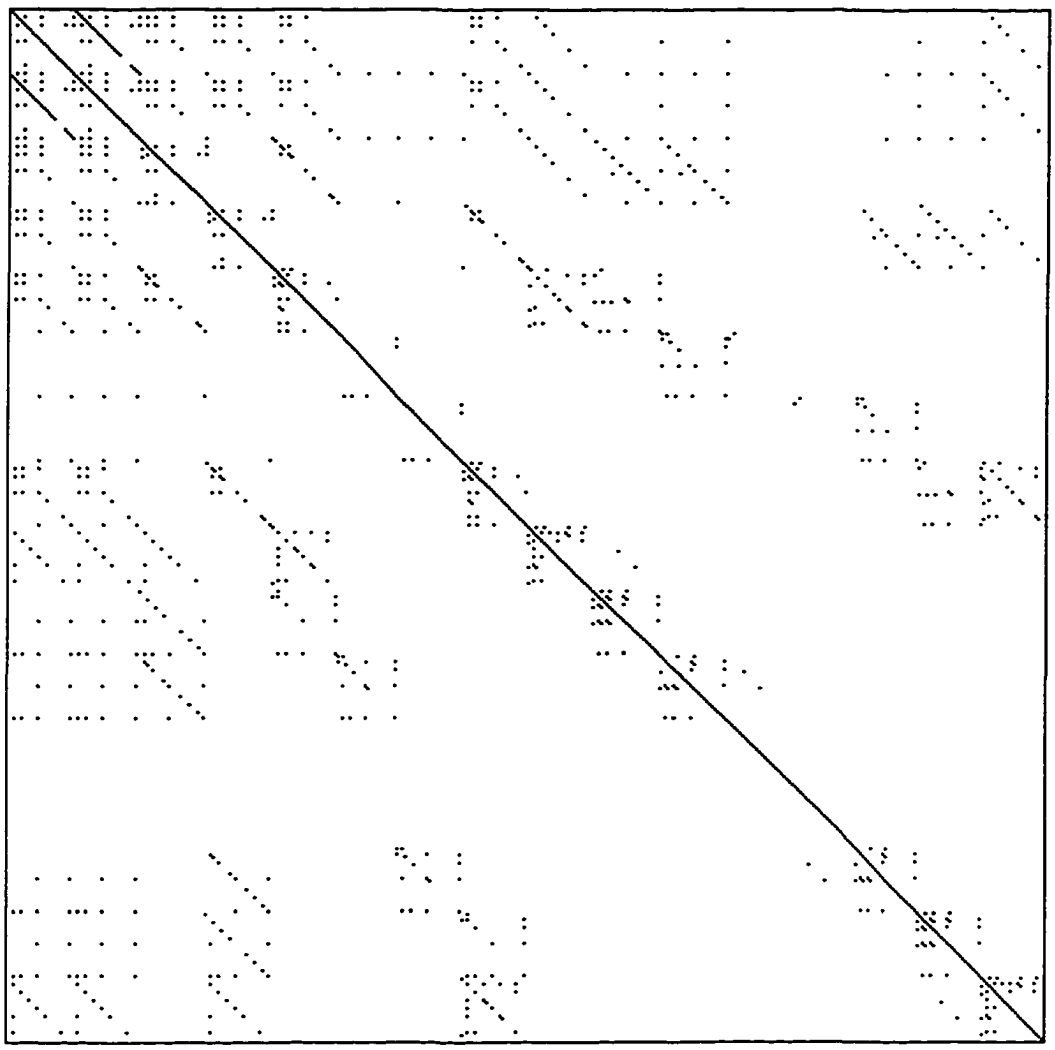


Fig. 8. The system matrix map after transformation by using the Haar wavelet and thresholded using a value of 0.05. Black spots denote nonzero elements.

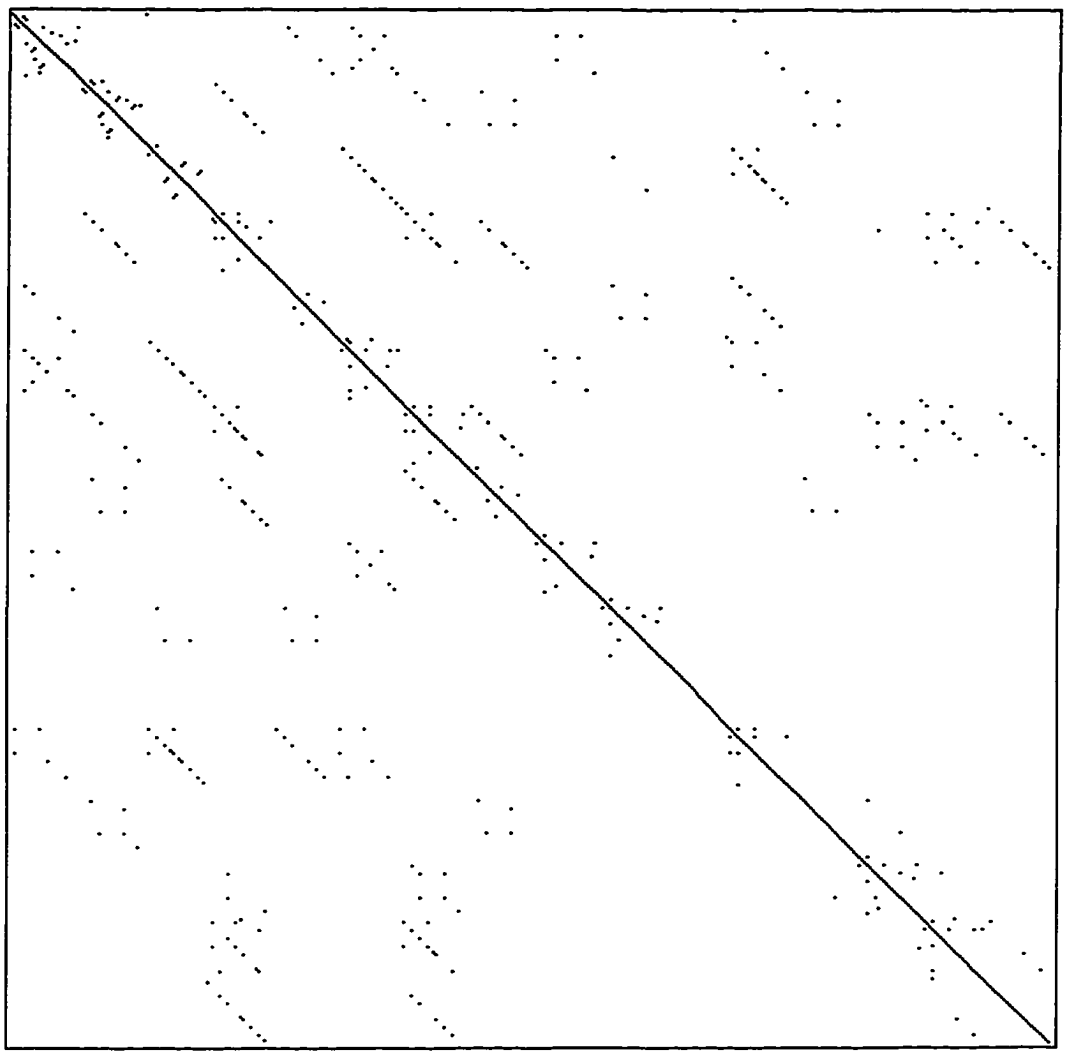


Fig. 9. The system matrix map after transformation by using DAUB4P and thresholded using a value of 0.05. Black spots denote nonzero elements.

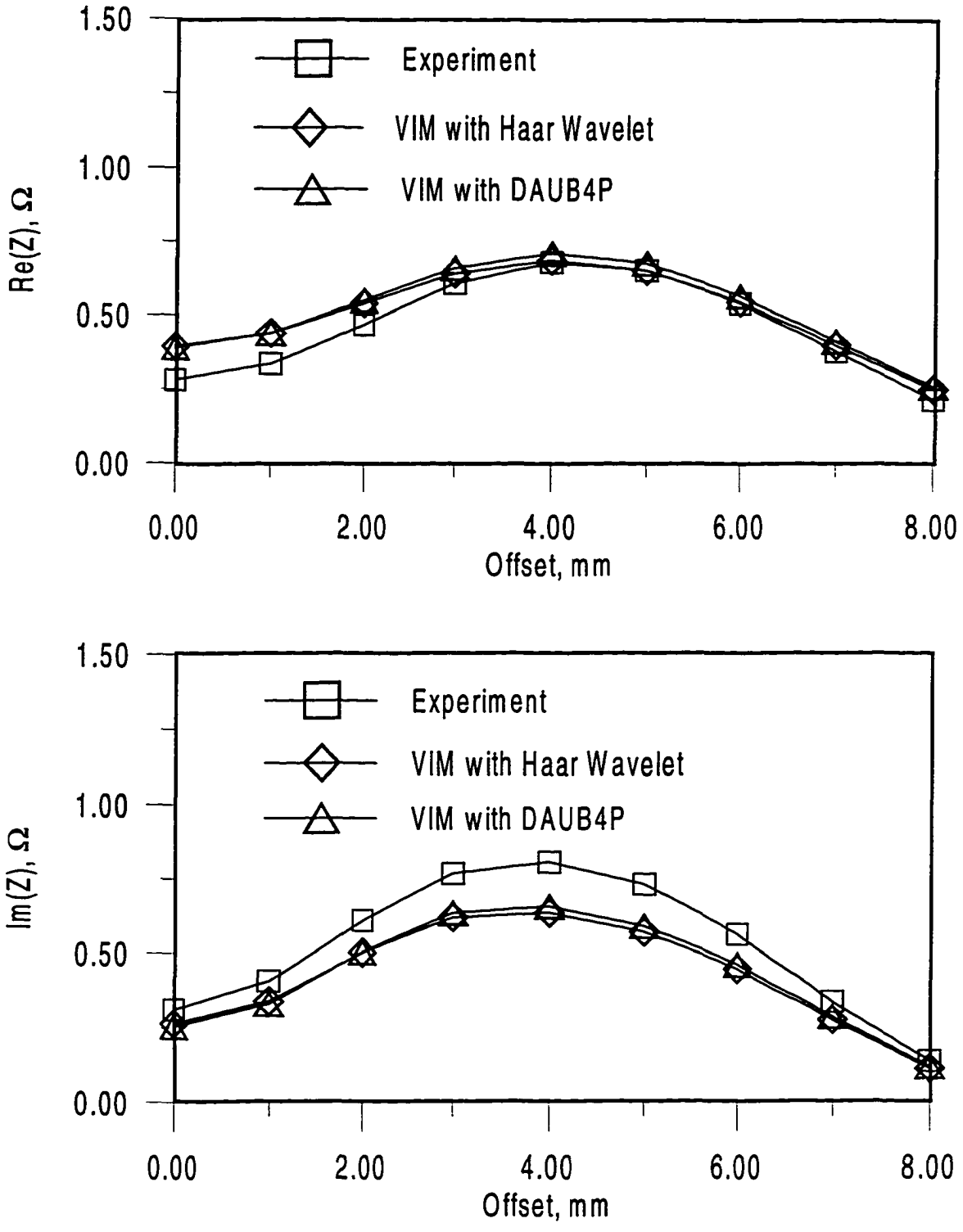


Fig. 10. Comparison of experimental results and the results obtained by using the Haar wavelet and DAUB4P with a threshold of 0.05.

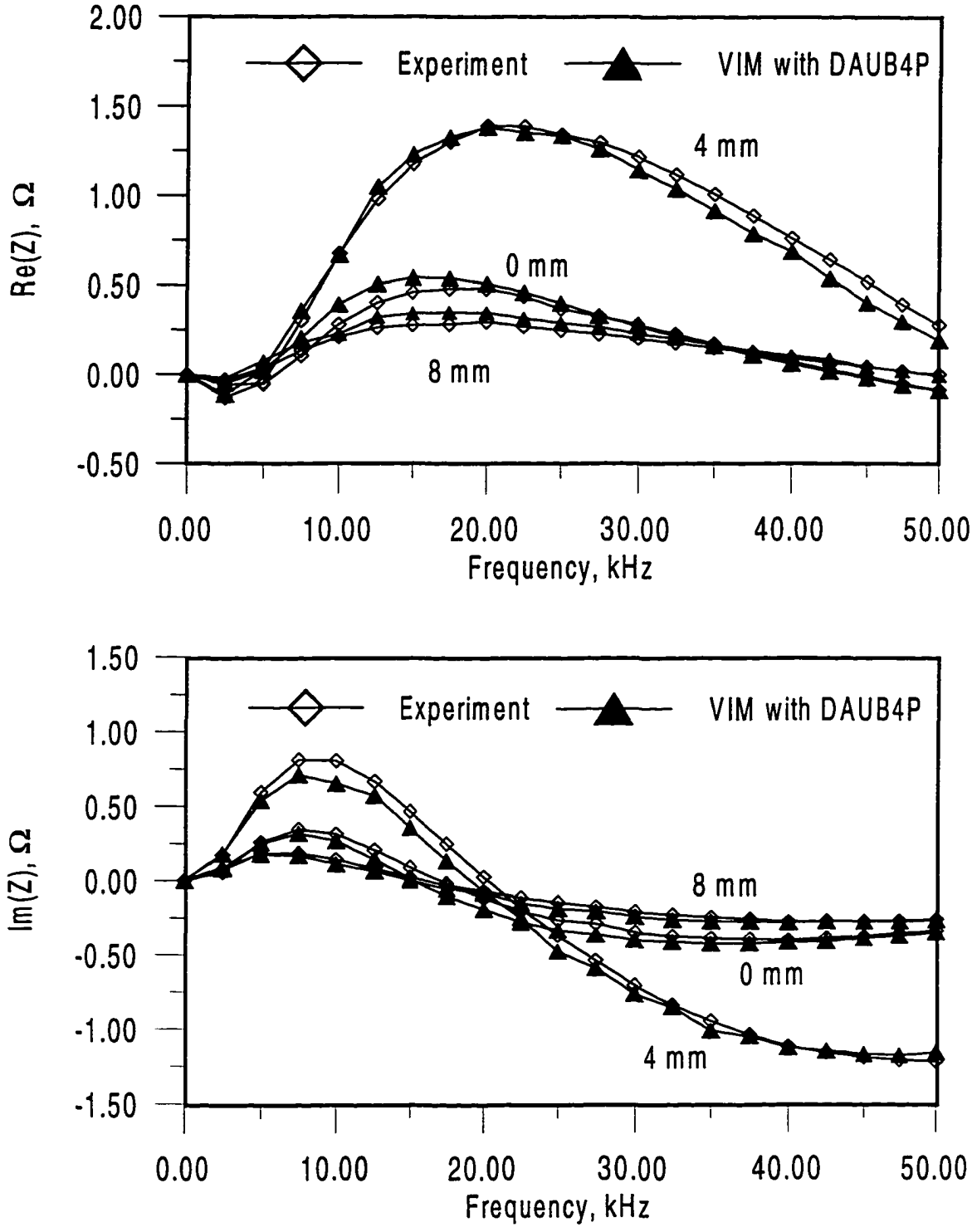


Fig. 11. Comparison of experimental results and the results obtained by using DAUB4P and a threshold of 0.005 for the frequency range from 2.5 kHz to 50 kHz.

APPENDIX B

NONLINEAR IMAGE RESTORATION USING RADIAL BASIS FUNCTION NETWORKS AND GENETIC ALGORITHMS

A paper to be submitted to
IEEE Transactions on Image Processing

Bing Wang, John P. Basart and John C. Moulder

Department of Electrical and Computer Engineering
and
Center for NDE
Iowa State University
Ames, Iowa 50011

Abstract

In this paper we present a nonlinear image restoration method based on radial basis function networks and genetic algorithms. Two-dimensional measurements made with nonlinear sensors, especially those involving diffusive fields, such as eddy current testing and thermal wave imaging, are usually a blurred and distorted version of the physical distribution to be measured. The underlying physical transformations in these sensors are usually smooth nonlinear mappings which generally can only be solved by numerical models. The image restoration problem for these applications is difficult because of the slow speed of available numerical models and the ill-posedness of the inverse mappings. Gradient based methods usually are not suitable because they tend to stick to local minima due to the ill-posedness of the problem. Combinatorial optimization methods, such as simulated annealing and genetic algorithms, are also difficult to apply because of the slow speed of the numerical models. In this paper we use radial basis function networks to approximate these forward mappings in a compact subset of the input space by learning from a set of training samples. The radial basis function networks can be made much faster than the numerical models so that we can use genetic algorithms to find the optimal restoration image which minimizes the forward prediction error. We discuss how to build the radial basis function network forward model and give the upper bound for the approximation error. We also discuss how to formulate the genetic algorithms to find the optimal restoration image. Application of this method to the eddy current image restoration problem is discussed and numerical restoration results are given for experimental measurement data from a surface-breaking crack on an thick aluminum plate.

1. Introduction

An image represents the distribution of a physical quantity within a certain region. To measure the distribution, we use sensors which transform one physical quantity to another physical quantity (usually electrical signals) so that the distribution can be easily measured and recorded. Depending on the physical process involved in the transformation, the sensor output can be a nonlinear function of the sensor input. Also, many sensors are not focused and they generate blurred representations of the physical distributions to be measured. Most of these blurring sensors can be modeled as a smooth nonlinear mapping from the input distribution to the output distribution. Typical examples include sensors involving diffusive fields, such as sensors used in eddy current testing and thermal wave imaging. In complex test geometries, the forward modeling problems of these sensors usually cannot be solved analytically and numerical models must be used. Since most numerical models require the inversion of a large system matrix, they have high computational complexity in both time and space.

To recover the original distribution from the blurred measurement, we need to develop image restoration methods for these applications. This nonlinear image restoration problem is more difficult than traditional nonlinear image restoration problems that can be found in photochemical or photoelectric systems. First of all, for this problem we do not have an analytical representation for the forward mapping. Therefore, it is very difficult to derive an analytical expression for the gradient of the restoration error. As a result, many image restoration methods that require knowledge of the gradient, such as the steepest descent method and the conjugate gradient method, cannot be used. Furthermore, the inverse mapping for these measurements is generally highly ill-posed due to averaging introduced in the forward mapping (changes in details of the input distribution lead to very small changes in the output distribution). Even if we can use gradient based methods, for these applications we will not be able to obtain good restoration results because gradient based methods can easily stick to local minima, and there are a lot of local minima in the error surface of an ill-posed problem. To find the global minima for an ill-posed problem, we can use combinatorial optimization methods, such as simulated annealing and genetic algorithms.

However, both simulated annealing and genetic algorithms are computationally intensive algorithms which require solving the forward problem many times. Due to the slow speed of the numerical models, it is not practical to apply these combinatorial optimization methods to the solution of the nonlinear image restoration problem if any numerical model is used.

To speed up the forward computation, we use radial basis function networks to approximate the forward mapping in a compact subset of the whole input space. The radial basis function network is an approximation neural network that can be used to find a surface in a multi-dimensional space that provides a best fit to training data obtained from a unknown function. Radial basis functions were first introduced in the solution of the real multivariate interpolation problem by Powell [1] and were later extended to the design of neural networks by Broomhead and Lowe [2]. Studies of the theory, design and application of the radial basis function network have been presented in [2, 3, 4]. Poggio and Girosi [3] developed regularization networks from approximation theory with the radial basis function network as a special case. It has been proven that a radial basis function neural network is a uniform approximator [4]. In this paper, we establish approximate forward models for the nonlinear image restoration problem using radial basis function networks. The training data set of the radial basis function network can be obtained from numerical models or from experimental measurements. We derived the upper bound of the approximation error. Guidelines on the selection of the training data set can be observed from this derivation. The radial basis function network based forward model can be made much faster than numerical models if the dimensionality of the subspace to be approximated is limited, and if the unknown forward mapping is sufficiently smooth.

Based on the fast approximate forward model, we developed genetic algorithms to solve the nonlinear image restoration problem. Genetic algorithms [5, 6] are a set of powerful heuristic search algorithms which can be used to solve large scale combinatorial optimization problems, including many that are NP-hard. The original development of genetic algorithms was done by Holland [7] in the field of artificial intelligence more than 20 years ago. But only recently have genetic algorithms been extensively studied and applied to many problems. Discussions on the concept and theory of genetic algorithms can be found in

Reference [8]. Applications of genetic algorithms include image reconstruction [9], solution of an electromagnetic inverse problem [10], optimal design of digital filter [11], and image segmentation [12], among many others. By formulating the nonlinear image restoration problem as a cost minimization problem, genetic algorithms can be used to find the optimal restoration image. We applied this method to eddy current image restoration and tested it using experimental data obtained from a crack on a thick aluminum plate. Numerical restoration results from the experimental data have shown that the nonlinear image restoration method developed in this study can provide better results than methods based on a linear approximation while keeping a reasonable execution time.

The rest of the paper is organized as follows: in Section 2 we give a formal definition of the nonlinear image restoration problem to be solved in this study; in Section 3 we give a brief review of the radial basis function network and discuss how to build the fast forward model using the radial basis function network. We also derive the upper bound of the approximation error and give guidelines on the selection of a training data set. In Section 4 we discuss how to use genetic algorithms to solve the nonlinear image restoration problem. We also discuss the encoding, fitness computation, selection, mutation and crossover operations of the genetic algorithms developed for the nonlinear image restoration problem. In Section 5 we discuss the application of the nonlinear image restoration method to the eddy current image restoration problem and present numerical results for both synthesized data and numerical data. The paper is summarized and concluded in Section 6.

2. The Nonlinear Image Restoration Problem

To measure physical quantities or distributions, we use various kinds of sensors. Sensors transform physical quantities or distributions into electrical signals which can be easily measured and recorded. The exact form of the transformation from the sensor input to the sensor output is dependent on the underlying physical process in the sensor transformation process. To measure a distribution, we want a sensor which is linear and focused so that the measurement is an accurate representation of the unknown distribution. Unfortunately, most sensors are nonlinear in nature and have limited resolution. Measurements made with these sensors are blurred and distorted representations of the unknown distributions. Nonlinear image restoration techniques must be used to restore the original distribution. In many applications, we have a certain amount of prior information about an unknown distribution and only need to consider the problem in a compact subset of the whole input space.

Consider the nonlinear mapping from a digital image of size p_1 by q_1 to another digital image of size p_2 by q_2. By transforming the two-dimensional images into one-dimensional vectors, we can change the original mapping to a map $s: R^M \rightarrow R^N$, where $M = p_1 q_1$ and $N = p_2 q_2$. This representation is also applicable to mappings between higher dimensional images. Based on our prior knowledge about an unknown distribution, we only need to consider input vectors that fall within a compact subset P of the whole input space R^M. We also define the set $Q = \{s(\mathbf{X}) | \mathbf{X} \in P\}$ as the corresponding output subset for the input subset P. In this paper we consider smooth nonlinear mappings whose underlying physical process may involve diffusive fields. We are interested in smooth mappings that are continuous and have continuous and bounded first derivatives in P. That is, if $\mathbf{X} = [x_1, \cdots, x_M]^T \in P$ and $\mathbf{Y} = [y_1, \cdots, y_N]^T \in Q$, then

$$0 \leq A_l \leq \left\| \frac{\partial Y}{\partial x_l} \right\| \leq B_l, \quad l = 1, 2, \cdots, M, \tag{1}$$

where $\|\cdot\cdot\|$ is the l_2 norm.

Assume $\mathbf{X}_0 \in P$ is the unknown distribution , $\mathbf{Y}_0 = s(\mathbf{X}_0) \in Q$ is the corresponding ideal sensor output. A noisy measurement of $\mathbf{Y}_0$ is

$$\hat{\mathbf{Y}} = \mathbf{Y}_0 + \mathbf{N}_0, \tag{2}$$

where $\mathbf{N}_0$ is a vector containing the noise image. We can define the nonlinear image restoration problem as to find a restoration $\hat{\mathbf{X}}$ of $\mathbf{X}_0$ such that the mean square restoration error

$$\varepsilon = E\left\{\left\|\hat{\mathbf{X}} - \mathbf{X}_0\right\|^2\right\} \tag{3}$$

is minimized. Based on this definition, we need to have knowledge about the inverse mapping $t = s^{-1}: R^N \rightarrow R^M$ and statistics on $\mathbf{X}_0$ and $\mathbf{N}_0$ to solve the minimization problem. However, in most applications, this information is unavailable or difficult to obtain. Alternatively, we can define the nonlinear image restoration problem as to find a restoration $\hat{\mathbf{X}}$ of $\mathbf{X}_0$ such that the forward prediction error $\left\|s(\hat{\mathbf{X}}) - \hat{\mathbf{Y}}\right\|$ is minimized. Due to the ill-posedness of the inverse mapping, minimization of the forward prediction error usually does not guarantee the minimization of the restoration error. To regularize the solution, we define the problem as to find a restoration $\hat{\mathbf{X}}$ of $\mathbf{X}_0$ to minimize the cost function

$$\varepsilon(\mathbf{X}) = \left\|s(\mathbf{X}) - \hat{\mathbf{Y}}\right\|^2 + \lambda \varepsilon_c(\mathbf{X}), \tag{4}$$

where λ is a regularization constant, and $\varepsilon_c(\mathbf{X})$ is a regularization function which represents our prior knowledge on the restoration. Generally this regularization function is a smoothness constraint which requires the restoration to have smooth transitions.

Due to the nonlinearity in the forward mapping, usually we need to use iterative methods to solve the minimization problem defined in equation (4). Many algorithms for finding the minima of a nonlinear cost function require knowledge of the gradient of the cost function. However, for the nonlinear image restoration problem defined in equation (4), we generally cannot find an analytical expression for the gradient because the forward problem can only be solved numerically. Therefore, gradient based methods are not applicable to the solution of this minimization problem. Combinatorial optimization algorithms, such as simulated annealing and genetic algorithms do not require knowledge about the gradient, but require solving the forward problem many times. Therefore, they are applicable only if a sufficiently fast forward model is available.

3. Approximation of the Forward Mapping Using a Radial Basis Function Neural Network

3.1 The Radial Basis Function Network

A radial basis function (RBF) network consists of an input layer, a hidden layer and an output layer. The input vector of the network is passed to the hidden layer nodes via unit connection weights. The input-output relationship of a hidden layer node is defined by a radial basis function. Generally the same type of radial basis functions are used for all the hidden layer nodes. Associated with each hidden layer node is a center $\mathbf{m}_i$ in the input space. The input to a radial basis function is the Euclidean distance between the input vector and the center $\mathbf{m}_i$. The output of a hidden layer node can be defined as

$$\phi_i(\mathbf{X}) = g\left(\left\|\mathbf{X} - \mathbf{m}_i\right\|\right), \tag{5}$$

where $\mathbf{X}$ is the input vector, and $g\left(\|\cdot\cdot\|\right)$ is the radial basis function. There are many different types of radial basis functions, and it has been shown, by theoretical investigations and practical results that the performance of the radial basis function network is not crucially dependent on the selection of the type of radial basis function [3]. However, as shown by

Light [4], there is a class of radial basis functions which gives a positive definite N by N interpolation matrix from a set of N distinctive input vectors. Since our objective is to use the radial basis function network to approximate an unknown smooth mapping, we only consider radial basis functions which belong to this class. As indicated by Light, two of the radial basis functions in this class are:

(i) *The Gaussion function:*

$$g(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right), \tag{6}$$

(ii) *The inverse multiquadric function:*

$$g(r) = \frac{1}{(r^2 + c^2)^{1/2}}, \tag{7}$$

where r, a non-negative number, is the distance from the input vector $\mathbf{X}$ to the radial basis function center $\mathbf{m}_i$, σ is the width of the Gaussian function, and c is the width of the inverse multiquadric function.

The output y_j of node j in the output layer is a linear combination of the outputs in the hidden layer

$$y_j = \sum_i w_{ji}\phi_i(\mathbf{X}), \tag{8}$$

where w_{ji} is the linear weight connecting node j in the output layer and node i in the hidden layer.

It has been proven that a radial basis function neural network is a uniform approximator [4]; it can approximate arbitrarily well any multivariate function on a compact domain if a sufficient number of radial basis function units are used. Therefore, it is suitable

for being used as an interpolator to approximate an unknown smoothing mapping based on a set of training samples.

3.2 Approximation of the Forward Mapping Using a Radial Basis Function Network

To solve the minimization problem in equation (4), we need to compute the forward mapping s many times since the inverse problem is nonlinear and some type of iterative method must be used to solve the problem. In many applications, computation of the forward mapping involves the solution of a partial differential equation (PDE) with certain boundary conditions. For complex test geometry, an analytical solution of the PDE is generally not available and numerical models must be used. Most numerical models divide the solution region into volume elements or surface elements and transform the PDE into a set of linear equations which can then be solved directly. The solution of a linear system involving a matrix inversion operation is basically an $O(M^3)$ operation, where M is the number of elements used in the forward model. For the nonlinear image restoration problem, the number of elements used in the model are usually on the order of thousands which leads to a long execution time. The slow execution speed of those numerical models becomes the major barrier to the applications of combinatorial optimization methods, such as simulated annealing and genetic algorithms, to the solution of the nonlinear image restoration problem.

To speed up the forward computation, we can use an RBF network to learn the forward mapping s in the input subset P based on a set of training samples. The training samples can be obtained by using a numerical model or by performing experimental measurements on calibration test samples. The benefit of using the RBF network is that we can reduce the computational complexity of the forward computation to a large extent if the dimensionality of the subset P is limited and the forward mapping is smooth enough. If the radial basis function network has M input layer nodes, L hidden layer nodes and N output layer nodes, then the complexity of the forward computation is $O(ML+NL)$, or $O(ML)$ if N and M are on the same order. The number of hidden layer nodes L necessary for approximating the forward mapping is dependent on the dimensionality of the subset P, the smoothness of the forward mapping, and the maximum approximation error allowed in the

subset P, as we shall see in the next section. If the forward mapping is smooth enough and if we can limit the dimensionality of P by our prior information about the unknown distribution, it is possible to bound L with $O(M)$. Thus, we obtain an approximation of the forward mapping in P with a complexity $O(M^2)$, which is one order faster than the numerical models. The one-order difference in computational complexity can be significant when the number of elements M is large, as in the nonlinear image restoration problem.

To build the fast forward model, we are interested in factorizable radial basis functions for the ease of implementation. The only factorizable radial basis function is the multivariate Gaussian function [3] which is given by

$$g_i(\mathbf{X}) = \exp\left[-\frac{1}{2}(\mathbf{X}-\mathbf{m}_i)\Sigma_i^{-1}(\mathbf{X}-\mathbf{m}_i)\right], \tag{9}$$

where $\mathbf{m}_i$ is the center of the Gaussian function, and Σ_i is the covariance matrix of the multivariate Gaussian function. If the covariance matrix is not diagonal, the function is factorizable only after a similarity transformation which transforms Σ_i to a diagonal form. For ease of implementation, we only consider the direct factorizable Gaussian function when the covariance matrix is a diagonal matrix, i.e., $\Sigma_i = diag\{\sigma_{i1}^2, \sigma_{i2}^2, \cdots, \sigma_{iM}^2\}$. In this case the multivariate Gaussian function can then be expressed as

$$g_i(\mathbf{X}) = \exp\left\{-\sum_{l=1}^{M}\frac{(x_l - m_{il})^2}{2\sigma_{il}^2}\right\}, \tag{10}$$

and the output y_j of node j in the output layer becomes

$$y_j = \sum_{i=1}^{L} w_{ji} \exp\left\{-\sum_{l=1}^{M}\frac{(x_l - m_{il})^2}{2\sigma_{il}^2}\right\}. \tag{11}$$

The training of the RBF network is to find the optimal values for $\mathbf{m}_i$, σ_{il}, w_{ij} so as to minimize the learning error. In general, a nonlinear optimization algorithm must be used for the training of the hidden layer parameters, and a linear optimization algorithm must be used for the output layer parameters. For our problem of approximating the smooth mapping s, we considered two cases:

(1) *Interpolation network*

In this case, we have a set of L training samples $[\mathbf{X}_i, \mathbf{Y}_i], i = 1, \cdots, L$ about the unknown smooth nonlinear mapping. We can use the training sample input vectors as the centers for the radial basis functions, i.e.,

$$\mathbf{m}_i = \mathbf{X}_i, \quad i = 1, \cdots, L, \tag{12}$$

and the RBF network becomes a nonlinear interpolator. Furthermore, we assume equal generalization capability in all the dimensions for each center, i.e., $\sigma_{il} = \sigma_0, i = 1, \cdots, L, l = 1, \cdots, M$. Based on this assumption, contours of each multivariate Gaussian function are hyperspheres around the center of the Gaussian function. To decide the optimal output layer weights, we need to solve

$$\tilde{\mathbf{Y}} = \mathbf{W}\tilde{\mathbf{G}}, \tag{13}$$

where $\tilde{\mathbf{Y}} = [\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_L]$ is an N by L matrix containing the desired output vectors given by the training samples, $\tilde{\mathbf{G}}$ is an L by L matrix containing the hidden layer outputs for the L input vectors, and $\mathbf{W}$ is an N by L matrix containing output layer weights. Since the multivariate Gaussian belongs to the class of radial basis function as defined by Light, the matrix $\tilde{\mathbf{G}}$ is positive definite and thus invertible. Then the weight matrix can be solved directly by

$$\mathbf{W} = \tilde{\mathbf{Y}}.\tilde{\mathbf{G}}^{-1} \tag{14}$$

(2) *Self-organizing of centers*

In case we have a large sample of training samples and there is redundancy in the training data set, we cannot use as many hidden layer nodes as the number of training samples because this will lead to a large size RBF network which is slow for the purpose of building a fast forward model. To reduce the number of centers, we use the K-Mean algorithm to find L cluster centers in W training samples with L much smaller than W. In this case, the optimal output layer weights cannot be computed directly because there are more equations than unknowns so that the problem is overdetermined. One way to compute the linear output layer weights is to use a Least Mean Square (LMS) algorithm to iteratively compute the optimal output layer weights [3]

$$w_{ji}(n+1) = w_{ji}(n) + \eta(y_j - o_j)g_i, \tag{15}$$

where w_{ji} is the weight between output layer node j and hidden layer node i, y_j is the desired output of output layer node j, o_j is the actual output of the output layer node j, g_i is the output of the hidden layer node i, and η is a learning rate parameter.

3.3 Bound of the Approximation Error

One important problem in using the RBF network to approximate the unknown smooth nonlinear mapping is to provide a bound on the approximation error in the subspace P. The knowledge of this error bound not only provides us with an estimate for the restoration error, but also leads to guidelines on the selections of the training data set. In this section we derive an expression for the bound of the approximation error for the interpolation network.

Denote the RBF network as a nonlinear map r: $R^M \rightarrow R^N$. For the interpolation network, we have $r(\mathbf{X}_i) = s(\mathbf{X}_i), i = 1, \cdots, L$. Given an arbitrary input vector $\mathbf{X} \in P$, the l_2

norm $\|r(\mathbf{X}) - s(\mathbf{X})\|$ defines the approximation error of the radial basis function network. We can define the approximation error as a function of $\mathbf{X}$

$$e(\mathbf{X}) = \|r(\mathbf{X}) - s(\mathbf{X})\|. \tag{16}$$

To find the bound on $e(\mathbf{X})$, let

$$\mathbf{X}_k = \left\{\mathbf{X}_i \middle| \|\mathbf{X} - \mathbf{X}_i\| \le \|\mathbf{X} - \mathbf{X}_j\|, i, j \in 1, \cdots, L \text{ and } j \ne i\right\} \tag{17}$$

be the training input vector that is closest to the input vector $\mathbf{X}$. We have

$$\begin{aligned} e(\mathbf{X}) &= \|r(\mathbf{X}) - s(\mathbf{X})\| \\ &= \|r(\mathbf{X}) - r(\mathbf{X}_k) + s(\mathbf{X}_k) - s(\mathbf{X})\| \\ &\le \|r(\mathbf{X}) - r(\mathbf{X}_k)\| + \|s(\mathbf{X}) - s(\mathbf{X}_k)\|. \end{aligned} \tag{18}$$

because $\mathbf{X} = [x_1, x_2, \cdots, x_M]^T$ and $\mathbf{X}_k = [x_{k1}, x_{k2}, \cdots, x_{kM}]^T$. The second term on the right in the above equation can be further expressed as

$$\begin{aligned} \|s(\mathbf{X}) - s(\mathbf{X}_k)\| &= \|s(x_1, x_2, \cdots, x_M) - s(x_{k1}, x_2, \cdots, x_M) + s(x_{k1}, x_2, \cdots, x_M) - s(x_{k1}, x_{k2}, \cdots, x_M) \\ &\quad + \cdots + s(x_{k1}, x_{k2}, \cdots, x_M) - s(x_{k1}, x_{k2}, \cdots, x_{kM})\| \\ &\le \|s(x_1, x_2, \cdots, x_M) - s(x_{k1}, x_2, \cdots, x_M)\| + \|s(x_{k1}, x_2, \cdots, x_M) - s(x_{k1}, x_{k2}, \cdots, x_M)\| \\ &\quad + \cdots + \|s(x_{k1}, x_{k2}, \cdots, x_M) - s(x_{k1}, x_{k2}, \cdots, x_{kM})\| \\ &\le B_1\left[|x_1 - x_{k1}| + B_2\|x_2 - x_{k2}\| + \cdots + B_M|x_M - x_{kM}|\right] \\ &= \sum_{l=1}^{M} B_l|x_l - x_{kl}|. \end{aligned} \tag{19}$$

where we have used the fact that from (1) $\left\|\dfrac{\partial s(\mathbf{X})}{\partial x_l}\right\| \le B_l, \quad l = 1, \cdots, M$. The first term on the right in equation (18) can be expressed in a matrix form as

$$\begin{aligned}\left\|r(\mathbf{X})-r(\mathbf{X}_k)\right\| &= \left\|\mathbf{W}\mathbf{G}(\mathbf{X})-\mathbf{W}\mathbf{G}(\mathbf{X}_k)\right\| \\ &= \left[\left(\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right)^T\mathbf{W}^T\mathbf{W}\left(\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right)\right]^{1/2},\end{aligned} \tag{20}$$

where $\mathbf{G}(\mathbf{X})=\left[g_1(\mathbf{X}), g_2(\mathbf{X}), \cdots, g_L(\mathbf{X})\right]^T$ is a vector containing the hidden layer outputs for the input vector $\mathbf{X}$, and $\mathbf{G}(\mathbf{X}_k)$ is a vector containing hidden layer outputs for the input vector $\mathbf{X}_k$. From the Rayleigh-Ritz theorem [13] and the fact that $\mathbf{W}^T\mathbf{W}$ is an L by L symmetric matrix, we have

$$\begin{aligned}&\left[\left(\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right)^T\mathbf{W}^T\mathbf{W}\left(\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right)\right]^{1/2} \\ &\quad\le |\lambda|_{\max}\left[\left(\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right)^T\left(\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right)\right]^{1/2} \\ &\quad= |\lambda|_{\max}\left\|\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right\|,\end{aligned} \tag{21}$$

where $|\lambda|_{\max}=\left\{\max_i|\lambda_i|\;\middle|\;\lambda_i \textit{ is an eigenvalue of } \mathbf{W}^T\mathbf{W}\right\}$ is the largest absolute value of all the eigenvalues of $\mathbf{W}^T\mathbf{W}$. Now if $\left|\frac{\partial g_i(\mathbf{X})}{\partial x_l}\right|\le C_l, i=1\cdots, L, l=1,\cdots, M$, similar to equation (19), we can prove that

$$\left\|\mathbf{G}(\mathbf{X})-\mathbf{G}(\mathbf{X}_k)\right\|\le\sum_{l=1}^{M}C_l|x_l-x_{kl}|. \tag{22}$$

For the factorizable Gaussian function given by equation (10), we have

$$\frac{\partial g_i(\mathbf{X})}{\partial x_l}=-\frac{\left(x_l-m_{il}\right)}{\sigma_{il}^2}g_i(\mathbf{X}), i=1,\cdots. L, l=1,\cdots, M. \tag{23}$$

For find the bound on $\left|\frac{\partial g_i(\mathbf{X})}{\partial x_l}\right|$, let

$$\frac{\partial}{\partial x_k}\left(\frac{\partial g_i(\mathbf{X})}{\partial x_l}\right)=0, k=1,\cdots, M. \tag{24}$$

Solution of the above equations give $x_k = m_{ik}$, $k=1,\cdots, M$ *and* $k \neq l$ and $x_l = m_{il} \pm \sigma_{il}$.

The bound on $\left|\frac{\partial g_i(\mathbf{X})}{\partial x_l}\right|$ is thus given by

$$\left|\frac{\partial g_i(\mathbf{X})}{\partial x_l}\right| \leq \frac{e^{-1/2}}{\sigma_{il}}, i=1,\cdots, L, l=1,\cdots, M. \tag{25}$$

Defining $\sigma_l^0 = \min_{i=1}^{L} \sigma_{il}$, we have

$$C_l = \frac{e^{-1/2}}{\sigma_l^0}. \tag{26}$$

Combine equations (18), (19), (20)-(22), and (26), we have the bound for the approximation error as

$$\begin{aligned} e(X) &\leq \sum_{l=1}^{M}\left(B_l + \frac{|\lambda|_{\max} e^{-1/2}}{\sigma_l^0}\right)\left|x_l - x_{kl}\right| \\ &\leq \left(B + \frac{|\lambda|_{\max} e^{-1/2}}{\sigma_0}\right)\left\|X - X_k\right\|_1, \end{aligned} \tag{27}$$

where $B = \max_{l=1}^{M} B_l$, $\sigma_0 = \min_{l=1}^{M} \sigma_l^0 = \min_{i,l} \sigma_{il}$, and $\|\cdot\cdot\|_1$ is the l_1 norm. For the interpolation problem (in which the centers of the Gaussian functions are equal to the training sample input vectors, and the contours of the Gaussian functions are hyperspheres), the bound given by equation (27) gives us two important guidelines on the selection of the training data set:

1. The error bound is proportional to the Manhattan distance between the input vector and the training data sample closest to it. This suggests a uniform grid for minimizing the maximum approximation error. Although it may be more efficient if we put more training samples where the derivatives of the unknown function are large and less samples where the derivatives are small, it is not possible to find estimates of the derivatives without dense sampling in the input space. Without detailed knowledge on the derivatives, we should uniformly distribute the training sample points in the subspace P.
2. A proper value for σ_0 should be chosen so that the approximation error can be minimized. As can be seen from equation (27), a very small σ_0 could lead to a large approximation error ($|\lambda|_{\max}$ is also related to σ_0) because in this case, the width of the Gaussian functions are not sufficient to cover the space between the centers. A large σ_0 could help reduce the approximation error as the reconstructed surface has smoother transition from one Gaussian function center to another Gaussian function center. However, a very large σ_0 can make matrix $\tilde{\mathbf{G}}$ in equation (13) close to singular, and the matrix inversion in (14) becomes numerically intractable. A reasonable selection for σ_0, as suggested in [14], is

$$\sigma_0 = \frac{d}{\sqrt{2L}}, \tag{28}$$

where $d = \max_{i,j} \|\mathbf{X}_i - \mathbf{X}_j\|$ is the maximum distance between training sample input vectors.

4. Genetic Algorithms for Nonlinear Image Restoration

In this section we discuss how to use genetic algorithms to solve the nonlinear image restoration problem defined in Section 2. Using the definition given in equation (4), the nonlinear image restoration problem can be considered as the minimization of the cost function with respect to the input image. This minimization problem can be considered as a combinatorial optimization problem which can be solved by using genetic algorithms. In this section we discuss various aspects of formulating genetic algorithms to solve the minimization problem defined in equation (4), including solution encoding, fitness function, selection scheme, mutation operation, and crossover operation.

4.1 Encoding of the Solutions

The first step of genetic algorithms is to encode the solutions of a problem in a form that can be easily manipulated. The encoding solutions are also called chromosomes in genetic algorithms. In the nonlinear image restoration problem, a solution is a digital image in two or higher dimensional space. After transforming the multi-dimensional digital image into an one-dimensional array, a candidate solution can be encoding as a vector $\mathbf{X} = [x_1, \cdots, x_M]^T \in P$, where $P \in R^M$ is the compact subset in which the solution is located. The elements $x_1, \cdots, x_M$ are real numbers whose value may be digitized. For an eight-bit image $x_1, \cdots, x_M \in \{0, 1, \cdots, 255\}$ and for a binary image $x_1, \cdots, x_M \in \{0, 1\}$. The major advantage of using the vector form representation is its ease of use and its generality as any digital image can be mapped to a one-dimensional vector. However, the vector representation does implicitly enforce the constraint that $\mathbf{X}$ must lie in the compact subset P. As a result, not all vectors of the form $\mathbf{X} = [x_1, \cdots, x_M]^T$ are valid solutions. This brings some complications to the definitions of the crossover operation and the mutation operation as we must generate valid solutions after applying these operations. To check the validity of a solution, we can define an indicator function as

$$I(\mathbf{X}) = \begin{cases} 1, & \text{if } \mathbf{X} \in P, \\ 0, & \text{otherwise}, \end{cases} \tag{29}$$

whose implementation is problem dependent.

4.2 Fitness Function

The fitness function defines how well an individual is able to survive in the environment. To solve a combinatorial optimization problem, the fitness function is related to the objective function to be optimized. For the nonlinear image restoration problem, the objective is to minimize the cost function defined in equation (4). Because the objective of a genetic algorithm is to maximize fitness, we can define the fitness function as

$$f(\mathbf{X}) = \begin{cases} \dfrac{1}{\varepsilon(\mathbf{X})}, & \text{if } \varepsilon(\mathbf{X}) \neq 0, \\ E, & \text{otherwise}, \end{cases} \tag{30}$$

where E is a very large number used to avoid overflow.

To regularize the ill-posed inverse problem, we need to define proper regularization functions to be incorporated in equation (4). The ill-posedness of the image restoration problem is closely related to values of the lower bounds $A_l, l = 1, \cdots, M$. If the values are close to 0, the problem is highly ill-posed and a large regularization constant λ must be used. The regularization function $\varepsilon_s(\mathbf{X})$ can have various forms dependent on the application and our prior information about the restoration result. In many cases the regularization function is a smoothing constraint based on a linear high-pass operator, i.e.

$$\varepsilon_f(\mathbf{X}) = \|\mathbf{HX}\|^2, \tag{31}$$

where $\mathbf{H}$ is a high pass operator. For two-dimensional images, a widely used high pass operator is the Laplacian operator [15], given by

$$\mathbf{L} = \begin{bmatrix} 0.00 & -0.25 & 0.00 \\ -0.25 & 1.00 & -0.25 \\ 0.00 & -0.25 & 0.00 \end{bmatrix}. \tag{32}$$

Another widely used smoothing constraint is the entropy function. Image processing applications based on the Maximum Entropy Method (MEM) can be found in Reference [16]. When the entropy function is used, the regularization function can be defined as

$$\varepsilon_s(\mathbf{X}) = \sum_{l=1}^{M} x_l \log x_l. \tag{33}$$

To avoid zero values, pixels in the restoration image are usually initialized using small positive numbers. The entropy function is also a smoothing constraints because the image with the maximum entropy is a constant image.

4.3 Selection

One widely used selection scheme is the proportionate selection scheme in which the number of offspring for a given candidate is proportional to its fitness. However, for the fitness function defined in equation (30), the proportionate selection scheme is not appropriate for the following two reasons. First, because of the noise in the impedance change measurement, the minimum cost has a nonzero value and the maximum fitness may not be much larger the minimum fitness. Second, because the fitness defined in equation (30) introduces a nonlinear scaling to the cost function, the fitness difference between a good solution and a bad solution can be very small compared with the base value of the fitness. If the proportionate selection scheme is used, the expected number of instances of the good solution after selection is only slightly larger than that of the bad solution. As a result, the convergence of the genetic algorithms can be very slow. One way to deal with this problem is to scale the fitness value for each individual before the proportionate selection to extend

the range of the fitness difference. However, scaling of the fitness function may introduce individuals with negative fitness which requires special handling. An alternative way to deal with the small differences in fitness is to use the rank-based selection scheme. In the rank-based selection, the expected number of offspring for a candidate solution is only dependent on its fitness rank, not the absolute value of its fitness. In this selection scheme, first the individuals are sorted according to their fitness values. Then given a probability P_r, the rank-based selection always tries to pick the highest fitness individual with the probability P_r. If this is not successful, the individual with the second highest fitness value is selected with the probability P_r, and so on until an individual is selected or the lowest fitness individual is reached. Since in the rank-based selection, the expected number of instances of an individual is only dependent on its fitness rank, the result of selection is not sensitive to the scale of the fitness difference.

4.4 Mutation

The mutation operation adds random perturbations to the chromosomes of the candidate solutions in the population. In genetic algorithms, mutation is a random localized search mechanism which has a secondary role and is mainly used for restoring lost genetic materials that may not be recovered by crossover. For the one-dimensional vector chromosome $\mathbf{X} = [x_1, \cdots, x_M]^T$, we can use a variation of the trade mutation proposed in Reference [17]. For each element in the vector, we mutate it with probability P_m. If the element x_i is to be mutated, it is replaced by a random number uniformly distributed in the set $D - \{x_i\}$, where D is the set of all possible values of an element in $\mathbf{X}$. To ensure that the mutated chromosome is still in the compact subset P, the indicator function is used to check its validity. If the indicator function returns 0, the mutation operation is repeated until a valid solution is generated.

The above validity check process may result in a large waste of computer time. To make the process more efficient, we can incorporate our prior information about P in the mutation operation so that the results of mutation are more likely to fall in P. We can

achieve this by using a mask $\mathbf{Q} = [q_1, q_2, \cdots, q_M]^T$, where $q_l \in \{0, 1\}, l = 1, \cdots, M$. If the subset P has a large support in the l th dimension, we set $q_l = 1$. Otherwise, we set $q_l = 0$. In the mutation operation, we mutate the l th element of $\mathbf{X}$ with a larger probability P_{m1} if $q_l = 1$ and mutate it with a smaller probability P_{m2} if $q_l = 0$. By using this two-level mutation scheme, fewer repetitions of the mutation operation are necessary before a valid mutation result can be generated.

4.5 Crossover

Crossover operation is the key search mechanism of genetic algorithms. In the crossover operation, individuals in the population are paired off and exchange of genetic material between the two parents takes place. For the vector form chromosome, several types of crossover operations, e.g., the single-point crossover, two-point crossover, and the uniform crossover, may be used [5]. Given two parents $\mathbf{X}_1 = [x_{11}, \cdots, x_{1M}]^T$ and $\mathbf{X}_2 = [x_{21}, \cdots, x_{2M}]^T$, the single-point crossover scheme generates a random integer k which is uniformly distributed from 1 to M and exchange the elements in the two parents starting from the position k. The two-point crossover scheme generates two random integers k_1 and k_2 that are uniformly distributed from 1 to M and exchange the elements in the two parents between the position k_1 and the position k_2. For a pair of candidate solutions, both the single-point crossover and the two-point crossover operation takes place with a probability P_c. In the uniform crossover scheme, the exchange of genetic materials is on an element by element basis. For each position in the vector representation, the elements of the two parents are exchanged with the probability P_c.

For the nonlinear image restoration problem, the uniform crossover scheme is preferred because it has no positional bias. The crossover result of the uniform crossover is not dependent on how the two or higher dimensional image is mapped to the one-dimensional vector. Again, to ensure that the offspring are still in the compact subset P, the indicator function is used to check their validity. The crossover operation must be repeated if at least one of the two offspring is not a valid solution. For a specific application, it may be possible

to define a more efficient crossover operation which only creates offspring in the subset P and thus eliminate the need of repeating the crossover operation.

5. Application to Eddy Current Image Restoration

In this section we discuss how to apply the general image restoration method discussed in Section 3 and Section 4 to a real-world application - the eddy current image restoration problem. First, we give a brief introduction to the eddy current nondestructive technique and the eddy current image restoration problem. Then we discuss how to use the radial basis function network to approximate an eddy current numerical model and how to apply genetic algorithms to solve the nonlinear restoration problem. We also present restoration results for experimental two-dimensional eddy current measurements obtained from surface-breaking cracks on a thin aluminum plate and compare them with the restoration results of Wiener filtering.

5.1 The Eddy Current Image Restoration Problem

Eddy current nondestructive evaluation (NDE) is one of the most important NDE techniques. In an eddy current measurement, a coil driven by sinusoidal current is placed on a metal testpiece and the impedance change of the coil is measured as the coil is moved around the testpiece. By observing the change in impedance as a function of coil position, we can detect and characterize defects in the testpiece. Since eddy current testing only requires one side access to the testpiece, it can be used in complex test geometry where many other NDE techniques can not be easily applied. Typical applications of eddy current testing includes the inspection of steam generator tubing in nuclear power plants and the inspection of surface structures of aircraft. More information on the eddy current NDE technique can be found in Reference [18]. Due to the fact that eddy current testing is governed by diffusive electromagnetic fields, a two-dimensional impedance change image is a blurred version of the corresponding flaw image. Also, the impedance change distribution is a nonlinear function of the conductivity distribution of the flaw. The impedance change of an air-core coil placed on a testpiece can be computed by using the following reciprocity formula [19]

$$\Delta Z = -\frac{1}{I^2}\int_{V'} \delta\sigma(\mathbf{r}')\mathbf{E}^0(\mathbf{r}')\cdot\mathbf{E}(\mathbf{r}')dv', \tag{34}$$

where I is the current in the coil, $\delta\sigma(\mathbf{r}')$ is the conductivity change caused by a defect, $\mathbf{E}^0(\mathbf{r})$ is the incident electric field, and $\mathbf{E}(\mathbf{r})$ is the total electric field. The incident field is not a function of the flaw conductivity distribution and it can be determined if the coil parameters and test geometry are given. The total field is a more complex function dependent on both the incident field, test parameters and the flaw conductivity distribution. The relationship between the incident field, the total field and the flaw conductivity distribution can be expressed in the following volume integral equation [19]

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) + \int_{V'} \delta\sigma(\mathbf{r}')\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{E}(\mathbf{r}')dv', \tag{35}$$

where $\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')$ is the Green's function for the test geometry. Since the total field is a function of both the incident field and the conductivity change, the impedance change ΔZ is a nonlinear function of the conductivity change $\delta\sigma(\mathbf{r}')$.

The eddy current image restoration problem is a typical example of the nonlinear image restoration problem defined in Section 2. First, because it is governed by diffusive fields, the forward mapping is smooth and the derivatives of the forward mapping are bounded in a compact domain. Second, the eddy current image restoration problem is highly ill-posed due to the limited resolution of eddy current testing (details of the flaw conductivity distribution get lost in the volume integration in equation (34)). Third, existing numerical models for eddy current forward modeling are computationally intensive which makes the direct application of combinatorial optimization methods impractical.

5.2 Solving the Eddy Current Image Restoration Problem Using Radial Basis Function Networks and Genetic Algorithms

For the eddy current image restoration problem, we are interested in recovering the flaw shape from the impedance change image. To reduce the dimensionality of the problem, we assume the flaw is a void or a uniform inclusion and the flaw has identical depth span all over the flaw region. Based on this assumption, we can represent the flaw with a binary image. We can use the radial basis function network discussed in Section 3 to learn the nonlinear mapping from the binary flaw image to the complex impedance change image from training samples obtained through numerical computations or experiments. To simplify the generation of the training data set, we require that there be only one flaw region in the flaw image. This defines a subset in the whole solution space. Only flaw images in this subset are valid solutions. With this restriction, we can form the training flaw set by using flaws with various shapes and sizes. The actual shapes and sizes of the flaws used in the training set are dependent on our prior information about the defect. For instance, if we know that the flaw is a crack, we can use rectangular flaws with different lengths and widths to build the training set. If we have very little information about the flaw shape, we may need to create a training set that contains flaws with different shapes and sizes. The number of training flaws sufficient for a good restoration is dependent on the dimensionality of the solution space and the desired restoration error. Also, because of the noise in the impedance change measurement, using a large number of training samples may not lead to a small restoration error. As a result of this, usually we do not need to build a training set with a large number of samples if we have sufficient prior information on the flaw shape. The interpolation network discussed in Section 3 is appropriate for building the eddy current forward model because the number of training samples we need to use is limited. The training of the network is also straightforward (equation (14)) and there is no learning error.

Since the candidate solutions of the problem are binary images and the subset P denotes binary images with only one flaw region, we need to modify the general genetic algorithms discussed in Section (4) for this specific application. In the following we

summarize the major steps for the binary encoded genetic algorithms used for the eddy current image restoration problem

Encoding of the solution

The two-dimensional binary flaw image can be mapped into a one-dimensional binary vector (a binary string). Then a candidate solution can be expressed as

$$\mathbf{X} = \left[x_1, x_2, \cdots, x_M\right]^T, x_i \in \{0, 1\}, i = 1, \cdots, M. \tag{36}$$

Because in the fast forward model we have assumed that there is only one flaw region in the flaw image, not every binary string given by equation (36) is a valid candidate solution. Instead of using the indicator function, we have defined the crossover operation and mutation operation to always generate valid solutions.

Fitness function

For the eddy current image restoration problem, the cost function is defined as

$$C(\mathbf{X}) = \sum_i \sum_j \left|\Delta Z_{ij} - \Delta Z_{ij}^0\right|^2 + \lambda m, \tag{37}$$

where ΔZ_{ij} is the impedance change corresponding to candidate solution $\mathbf{X}$, ΔZ_{ij}^0 is the measured impedance change, λ is the regularization constant, and m is the number of directional changes in the flaw boundary (number of corners in the flaw). The first term in equation (37) gives the forward prediction error, and the second term forces the solution to have a regular boundary. The impedance change prediction ΔZ_{ij} is obtained using the radial basis function network.

Selection

The rank-based selection discussed in Section 4.3 is used.

Mutation

To generate a valid solution, we define the mutation operation as to randomly increase or decrease the flaw region by one pixel. If an increase is chosen, we find the set of pixels that are neighbor pixels of any of the boundary pixels in the candidate solution to be mutated. Then we randomly select one pixel from this set and add this pixel to the flaw region. On the other hand, if a decrease is chosen, we randomly pick a boundary pixel in the candidate solution and remove it from the flaw region. Since removing a boundary pixel that may separate the flaw region into two unconnected flaw regions, we need to repeat the random selection if the selected boundary pixel cannot be removed due to this reason.

Crossover

As discussed above, the major problem in defining the crossover operation is to ensure it generates valid offspring. In the eddy current image restoration problem, a valid solution can only have one flaw region. Let the two parent solutions be $\mathbf{X}_1$ and $\mathbf{X}_2$, and $\mathbf{I} = \mathbf{X}_1 \cap \mathbf{X}_2$, the following algorithm creates a valid offspring $\mathbf{Y}$:

1. If $\mathbf{I} = \varnothing$ (no intersection region), $\mathbf{Y} = \mathbf{X}_1$ with probability 0.5 and $\mathbf{Y} = \mathbf{X}_2$ with probability 0.5.
2. If $\mathbf{I}$ has only one flaw region, then the offspring are generated in the following way
   (1) Set $\mathbf{Y} = \mathbf{I}$.
   (2) While $A(\mathbf{Y}) < \frac{1}{2}\left(A(\mathbf{X}_1) + A(\mathbf{X}_2)\right)$, where $A(\mathbf{X})$ is a function which returns the flaw area of a flaw image $\mathbf{X}$,
      a. Let $\mathbf{B}$ be the set of pixels that are neighboring pixels of any of the boundary pixels in $\mathbf{Y}$, $\mathbf{B}_1 = \mathbf{B} \cap \mathbf{X}_1$, and $\mathbf{B}_2 = \mathbf{B} \cap \mathbf{X}_2$.
      b. $\mathbf{Y} = \mathbf{Y} \cup \mathbf{B}_1$ with probability 0.5 and $\mathbf{Y} = \mathbf{Y} \cup \mathbf{B}_2$ with probability 0.5.

3. If $\mathbf{I}$ has more than one flaw region, $\mathbf{Y} = \mathbf{X}_1$ with probability 0.5 and $\mathbf{Y} = \mathbf{X}_2$ with probability 0.5.

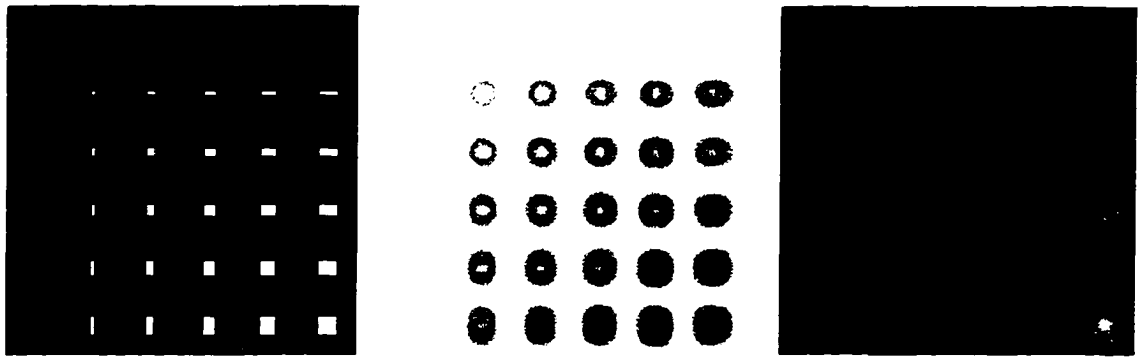
In the above crossover scheme, case 1 and case 3 are situations where a valid crossover operation is difficult to define. They are simply ignored because of their relatively rare occurrences. In case 2, we initialize the offspring with the intersection of the flaw regions of the two parents (where common genetic material is located). Then we randomly grow the offspring in the union of the flaw regions of the two parents until its flaw region reaches a reasonable size. This crossover scheme allows efficient exchange of genetic materials while keeping the offspring a valid candidate solution.

5.3 Numerical Results

To evaluate the performance of the nonlinear image restoration method on the eddy current image restoration problem, we tested it using experimental data obtained from a two-dimensional impedance change measurement for a surface-breaking crack on a thick aluminum plate. The crack has a length of 4.04 mm, a width of 0.21 mm, and a depth of 0.916 mm. The aluminum plate is thick enough to be considered as a half space. The coil used has an inner radius of 3.8 mm, an outer radius of 5.63 mm, and a height of 2.64 mm. A two dimensional scan was made on a 64 by 64 grid, with a grid separation of 0.4 mm. Before restoration, the data were preprocessed using polynomial background removal and spline smoothing.

To generate the training data set for the neural network forward model, we created 36 artificial flaws with various lengths and widths. The impedance change images for those artificial flaws were then computed by using a volume integral model with wavelet expansion [19]. The training flaw images and the real and imaginary impedance change images are shown in Fig. 1.

After the training data set was created, a radial basis function network with 4096 input layer nodes, 36 hidden layer nodes, and 8192 output layer nodes (4096 nodes for the real part impedance change and 4096 nodes for the imaginary part impedance change) was



(a) Flaw images (b) Re(Z) images (c) Im(Z) images

Fig. 1. The training data set for creating the neural network forward model.

used to approximate the nonlinear forward model based on the training data set. The interpolation network was used and the variances for the Gaussian functions were decided using equation (28). The output layer weights were decided using equation (14).

After the fast forward model was generated, a restoration test was performed using the genetic algorithm based restoration method. The test was performed with a crossover probability of 0.9, a mutation probability of 0.7, a probability for rank-based selection of 0.3, a population size of 10, and a maximum number of generations of 10. The real part and imaginary part impedance change images of the laboratory data and the restoration result from the genetic algorithm are shown in Fig. 2. As a comparison, we also give the restoration result from Wiener filter in Fig. 2. The Wiener filter used for this problem is obtained by using a linear approximate forward model for the impedance change.

(a) Re(Z) (b) Im(Z) (c) Result of GA (d) Result of Wiener filter

Fig. 2. Test results for the experimental data using the genetic algorithm (GA) based method and Wiener filter.

In the restoration results from both the genetic algorithm based method and the Wiener filter method, the restored flaw is shorter than the actual flaw. This is thought to be a result of the background noise in the laboratory data. From Fig. 2 we can see that the restoration result from the genetic algorithm based method is better than that of Wiener filter in terms of resolution and signal-to-noise ratio. Another advantage of the genetic algorithm based method is that in its restoration result there is no distortion due to a linear approximation, which can usually be found in the results of linear methods for relatively long cracks. However, the genetic algorithm based approach is more computationally intensive than the Wiener filter approach which makes it only applicable to a limited amount of data.

6. Conclusions

In this paper we have presented a nonlinear image restoration method based on radial basis function networks and genetic algorithms. This method is proposed to solve the image restoration problem for two-dimensional measurements made with nonlinear sensors involving complex physical phenomenon. The image restoration problem for these applications are difficult because of the slow speed of available numerical models and the ill-posedness of the inverse mappings. To overcome the difficulties associated with gradient based methods and direct combinatorial optimization methods, we have applied radial basis function networks to approximate these forward mappings in a compact subset of the input space by learning from a set of training samples. The radial basis function networks can be made much faster than the numerical models so that we can use genetic algorithms to find the optimal restoration image which minimizes the forward prediction error. We derived an upper bound for the approximation error and gave some guidelines on the selection of training data set. Test results on the experimental data from a surface-breaking crack on an aluminum plate showed the practicality of the method in the eddy current image restoration problem and its superior performance over the Wiener filter method. However, even after using the radial basis function network to speed up the forward model, the genetic algorithm based method still has higher computational complexity than most linear methods. Future work is desired to improve the convergence performance of the genetic algorithm based nonlinear image restoration method and to further reduce its execution time.

References

[1] M. J. D. Powell, "Radial basis functions for multivariable interpolation: A review", in *IMA Conference on Algorithms for the Approximation of Functions and Data.*, pp.143-167, RMCS, Shrivenham, UK, 1985.

[2] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks", *Complex Systems*, Vol. 2, pp. 321-355, 1988.

[3] T. Poggio and F. Girosi, "Networks for approximation and learning", *Proceedings of the IEEE*, Vol. 78, pp. 1481-1497, 1990.

[4] W. A. Light, "Some aspects of radial basis function approximation", in *Approximation Theory, Spline Functions and Applications* (S. P. Singh, ed.), NATO ASI Series, Vol. 256, pp. 163-190, Kluwer Academic Publishers, Boston, MA, 1992.

[5] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey", *Computer*, Vol. 27, pp. 17-26, 1994.

[6] K. Gallagher and M. Sambridge, "Genetic algorithms: A powerful tool for large-scale nonlinear optimization problems", *Computer & Geoscience*, Vol. 20, pp. 1229-1236, 1994.

[7] J. H. Holland, *Adaptation in Natural and Artificial Systems*, Univ. of Michigan Press, Ann Arbor, Michigan, 1975.

[8] L. Davis, ed., *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, 1991.

[9] C. C. Chiu, P. T. Liu, "Image reconstruction of a perfectly conducting cylinder by the genetic algorithm", *IEE Proc. Microwaves, Antennas and Propagat.*, Vol. 143, pp. 249-253, 1996.

[10] M. Kishimoto, K. Sakasai, K. Ara, "Solution of electromagnetic inverse problem using combinational method of Hopfield neural network and genetic algorithm", *J. Appl. Phys.*, Vol. 79, pp. 1-7, 1996.

[11] D. J. Xu, M. L. Daley, "Design of optimal digital filter using a parallel genetic algorithm", *IEEE Trans. Circuits and Systems, Part II, Analog and Digital Signal Processing*, Vol. 42, pp. 673-675, 1995.

[12] B. Bhanu, S. Lee, S. Das, "Adaptive image segmentation using genetic and hybrid methods", *IEEE Trans. Aerospace and Electronic Systems*, Vol. 31, pp. 1268-1291, 1995.

[13] R. A. Horn and c. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[14] S. Haykin, *Neural Network: A Comprehensive Foundation*, Chapter 7, Macmillan College Publishing, New York, 1994.

[15] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Chapter 5, Addison-Wesley, 1992.

[16] S. F. Gull and J. Skilling, "Maximum entropy method in image processing", *IEE Proceedings*, Vol. 131, Pt. F, pp. 646-659, 1984.

[17] C. B. Lucasius and G. Kateman, "Towards solving subset selection problems with the aid of the genetic algorithms", in *Parallel problem solving from Nature* (R. Manner and B. Manderick, ed.), Elsevier Science Publishers, Amsterdam, 1992.

[18] D. J. Hagemaier, *Fundamentals of Eddy Current Testing*, ASNT, 1990.

[19] B. Wang, J. C. Moulder, and J. P. Basart, "Wavelets in the solution of the volume integral equation: Application to eddy current modeling", accepted by *J. Appl. Phys.*

REFERENCES

[1] B. A. Auld, "Theoretical Characterization and Comparison of Resonant Probe Microwave Eddy-current Methods", in *Eddy Current Characterization of Material and Structures*, edited by G. Birnbaum and G. Free, ASTM, Philadelphia, pp. 332-347, 1981.

[2] W. S. Dunbar, "The Volume Integral Method of Eddy Current Modeling", *J. Nondestruct. Eval.*, Vol. 5, pp. 9-14, 1985.

[3] A. H. Kahn, "Boundary Integral Equation Methods for Two-Dimensional Models of Crack-Field Interactions", *J. Nondestruct. Eval.*, Vol. 7, pp. 3-14, 1988.

[4] N. Nakagawa and J. Chao, "Extended Magnetic Potential Method for Quasistatic Electromagnetism and Eddy Current Phenomena", in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, Plenum, New York, Vol. 15, 1996.

[5] W. S. Dunbar, "The Volume Integral Method of Eddy Current Modeling: Verification", *J. Nondestruct. Eval.*, Vol. 7, pp. 43-54, 1988.

[6] D. M. Mckirdy, "Recent Improvements to the Application of the Volume Integral Method of Eddy Current Modeling", *J. Nondestruct. Eval.*, Vol. 8, pp. 45-52, 1989.

[7] P. C. French and L. J. Bond, "Finite-Element Modeling of Eddy-Current Nondestructive Evaluation (NDE)", *J. Nondestruct. Eval.*, Vol. 8, pp. 55-71, 1988.

[8] M. L. Burrows, *A Theory of Eddy Current Flaw Detection*, Ph.D. thesis, University of Michigan, Ann Arbor, 1964.

[9] A. J. Bahr and B. A. Auld, "An Electromagnetic Model for Eddy-Current Imaging", *J. Nondestruct. Eval.*, Vol. 7, pp. 71-77, 1988.

[10] R. E. Beissner, "Slots vs. Cracks in Eddy Current NDE", *J. Nondestruct. Eval.*, Vol. 13, pp. 175-183, 1994.

[11] S. M. Nair and J. H. Rose, "Low-frequency Asymptotics for Eddy Currents in a Conducting Half-space in the Absence and Presence of Inhomogeneities", *J. Appl. Phys.*, Vol. 70, pp. 1924-1937, 1991.

[12] R. Satveli, J. C. Moulder, B. Wang, and J. H. Rose, "Impedance of a Coil Near an Imperfectly Layered Metal Structure: The Layer Approximation", *J. Appl. Phys.*, Vol. 79, pp. 2811-2821, 1996.

[13] E. Uzal, J. C. Moulder, S. Mitra, and J. H. Rose, "Impedance of Coils Over Layered Metals With Continuously Variable Conductivity and Permeability: Theory and Experiment", *J. Appl. Phys.*, Vol. 74, pp. 2076-2089, 1993.

[14] B. R. Groshong, G. L. Bilbro, and W. E. Snyder, "An Eddy Current Transducer Model for Image Restoration", *J. Nondestruct. Eval.*, Vol. 10, pp. 55-62, 1991.

[15] D. J. Hagemaier, *Fundamentals of Eddy Current Testing,* ASNT, Columbus, OH, 1990.

[16] P. Craven and G. Wahba, "Smoothing Noisy Data with Spline Functions", *Numer. Math.*, Vol. 31, pp. 377-403, 1979.

[17] B. A. Auld, S. R. Jefferies, and J. C. Moulder, "Eddy-Current Signal Analysis and Inversion for Semielliptical Surface Cracks", *J. Nondestruct. Eval.*, Vol. 7, pp. 79-94, 1988.

[18] L. Udpa and S. S. Udpa, "Solution of Inverse Problems in Eddy-Current Nondestructive Evaluation (NDE)", *J. Nondestruct. Eval.*, Vol. 7, pp. 111-120, 1988.

[19] J. M. Mann, L. W. Schmerr, and J. C. Moulder, "Neural Network Inversion of Uniform-Field Eddy Current Data", *Material Evaluation*, January 1991, pp. 34-39.

[20] J. H. Rose and S. M. Nair, "Exact Recovery of the DC Electrical Conductivity of a Layered Solid", *Inverse Problems*, Vol. 7, pp. L31-L36, 1991.

[21] J. C. Moulder, E. Uzal, and J. H. Rose, "Thickness and Conductivity of Metallic Layers from Eddy Current Measurements", *Rev. Sci. Instrum.*, Vol. 63, pp. 3455-3465, 1992.

[22] S. K. Burke, "Eddy-Current Inversion in the Thin-Skin Limit: Determination of Depth and Opening for a Long Crack", *J. Appl. Phys.*, Vol. 76, pp. 3072-3080, 1994.

[23] S. M. Nair and J. H. Rose, "Reconstruction of Three-dimensional Conductivity Variations from Eddy Current Data", *Inverse Problems*, Vol. 6, pp. 1007-1030, 1990.

[24] E. Uzal, J. C. Moulder, and J. H. Rose, "Experimental Determination of the Near-Surface Conductivity Profiles of Metals from Electromagnetic Induction (Eddy Current) Measurements", *Inverse Problems*, Vol. 10, pp. 753-764, 1994.

[25] L. D. Sabbagh and H. A. Sabbagh, "Eddy Current Modeling and Flaw Reconstruction", *J. Nondestruct. Eval.*, Vol. 7, pp. 95-110, 1988.

[26] B. R. Groshong, G. L. Bilbro, and W. E. Snyder, "Eddy Current Image Restoration by Constrained Gradient Descent", *J. Nondestruct. Eval.*, Vol. 10, pp. 127-137, 1991.

[27] S. J. Norton and J. R. Bowler, "Theory of Eddy Current Inversion", *J. Appl. Phys.*, Vol. 73, pp. 501-512, 1993.

[28] C. V. Dodd and W. E. Deeds, "Analytical Solutions to Eddy-Current Probe-Coil Problems", *J. Appl. Phys.*, Vol. 39, pp. 2829-2838, 1968.

[29] A. P. Raiche and J. H. Coggon, "Analytic Green's Tensors for the Integral Equation Modeling", *Geophys. J. R. Astr. Soc.*, Vol. 42, pp. 1035-1038, 1975.

[30] R. E. Beissner, "Analytical Green's Dyads for An Electrically Conducting Half-Space", *J. Appl. Phys.*, Vol. 60, pp. 855-858, 1986.

[31] J. T. Weaver, "The Quasi-Static Field of an Electric Dipole Embedded in a Two-Layer Conducting Half-Space", *Can. J. Phys.*, Vol. 45, pp. 1981-2002, 1967.

[32] V. Magni, G. Cerullo, and S. D. Silvestri, "High-accuracy Fast Hankel Transform for Optical Beam Propagation", *J. Opt. Soc. Am. A*, Vol. 9, pp. 2031-2033, 1992.

[33] J. Nachamkin and D. C. Hoffman, "Rapid Evaluation of Integer-Order Bessel Functions Using an FFT Algorithm", *IEEE Antennas and Propagation Magazine*, Vol. 33, pp. 33-35, 1991.

[34] W. L. Anderson, "A Hybrid Fast Hankel Transform Algorithm for Electromagnetic Modeling", *Geophysics*, Vol. 54, pp. 263-266, 1989.

[35] E. W. Hansen, "Fast Hankel Transform Algorithm", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-33, pp. 666-671, 1985.

[36] B. W. Suter, "Fast Nth-Order Hankel Transform Algorithm", *IEEE Trans. Signal Processing*, Vol. 39, pp. 532-536, 1991.

[37] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1992.

[38] L. Y. Luke, *Integrals of Bessel Functions*, McGraw Hill, New York, 1962.

[39] B. Wang, J. P. Basart, and J. C. Moulder, "Wavelet Expansion in Volume Integral Method of Eddy-Current Modeling", in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, Plenum, New York, Vol. 15, 1996.

[40] B. Wang, J. C. Moulder, and J. P. Basart, "Wavelets in the Solution of the Volume Integral Equation: Application to Eddy Current Modeling", accepted by *J. Appl. Phys.*

[41] Y. T. Chan, *Wavelet Basics*, Kluwer Academic Publishers, Boston, 1995.

[42] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[43] C. K. Chui, *An Introduction to Wavelets*, Academic Press, New York, 1992.

[44] M. Vetterli, "Wavelets and Filter Banks: Theory and Design", *IEEE Trans. on Signal Processing*, Vol. 40, pp. 2207-2232, 1992.

[45] C. E. Heil and D. F. Walnut, "Continuous and Discrete Wavelet Transforms", *SIAM Review,* Vol. 31, pp. 628-666, 1989.

[46] 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, 1989.

[47] B. Z. Steinberg and Y.Leviatan, "On the Use of Wavelet Expansion in the Method of Moments", *IEEE Trans. Antennas Propagat.*, Vol. 41, No. 5, pp. 610-619, 1993.

[48] J. C. Goswami, A. K. Chan, and C. K. Chui, "On Solving First-Kind Integral Equations Using Wavelets on a Bounded Interval", *IEEE Trans. Antennas Propagat.*, Vol. 43, pp. 614-622, 1995.

[49] G. Wang, "A Hybrid Wavelet Expansion and Boundary Element Analysis of Electromagnetic Scattering from Conducting Objects", *IEEE Trans. Antennas Propagat.* Vol. 43, pp. 170-178, 1995.

[50] B. Wang, J. P. Basart, and J. C. Moulder, "Fast Eddy Current Forward Models Using Artificial Neural Networks", in *Review of Progress in Quantitative Nondestructive Evaluation*, Edited by D. O. Thompson and D. E. Chimenti, Plenum, New York, Vol. 16, 1997.

[51] S. Haykin, *Neural Networks: a Comprehensive Foundation*, Macmilan, New York, 1994.

[52] S. Ergezinger and E. Thomsen, "An Accelerated Learning Algorithm for Multilayer Perceptrons: Optimization Layers by Layers", *IEEE Trans. Neural Networks*, Vol. 6, pp. 31-42, 1995.

[53] J. Zhang, G. G. Walter, Y. Miao, and W. N. W. Lee, "Wavelet Neural Networks for Function Learning", *IEEE Trans. Signal Processing*, Vol. 43, pp. 1485-1497, 1995.

[54] M. J. D. Powell, "Radial Basis Functions for Multivariable Interpolation: a Review", in *Algorithms for Approximation*, edited by J. C. Mason and M. G. Cox, pp. 143-167, Oxford, 1987.

[55] V. Kurkova, "Approximation of Functions by Perceptron Networks with Bounded Number of Hidden Units", *Neural Networks*, Vol. 8, pp. 745-750, 1995.

[56] S. Elanayar V.T. and Y. C. Shin, "Radial Basis Function Neural Network for Approximation and Estimation of Nonlinear Stochastic Dynamic Systems", *IEEE Trans. Neural Networks*, Vol. 5, pp. 594-603, 1994.

[57] J. A. Leonard, M. A. Kramer, and L. H. Ungar, "Using Radial Basis Functions to Approximate a Function and Its Error Bounds", *IEEE Trans. Neural Networks*, Vol. 3, pp. 624-627, 1992.

[58] M. T. Musavi, W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels, "On the Training of Radial Basis Function Classifiers", *Neural Networks*, Vol. 5, pp. 595-603, 1992.

[59] F. Girosi, "Some Extensions of Radial Basis Functions and Their Applications in Artificial Intelligence", *Computers Math. Applic.*, Vol. 24, No. 12, pp. 61-80, 1992.

[60] N. Arad, N. Dyn, D. Reisfeld, and Y. Yeshurun, "Image Warping by Radial Basis Functions: Application to Facial Expressions", *Graphical Models and Image Processing*, Vol. 56, pp. 161-172, 1994.

[61] J. Moody and C. Darken, "Fast-learning in Networks of Locally-tuned Processing Units," *Neural Comput.*, Vol. 1, pp. 281-294, 1989.

[62] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks", *IEEE Trans. Neural Networks*, Vol. 2, pp. 302-309, 1991.

[63] B. Wang, J. P. Basart and J. C. Moulder, *Signal and Image Processing Techniques for Two-dimensional Eddy Current Inspection*, Research Report, Center for NDE, Iowa State University, 1996.

[64] E. R. Doering and J. P. Basart, "Trend Removal in X-Ray Images", in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, Plenum, New York, Vol. 7, pp. 785-793, 1988.

[65] E. R. Dougherty, *An introduction to nonlinear image processing*, SPIE, Bellingham, WA, 1994.

[66] E. Oja, "A Simplified Neuron Model as a Principal Component Analyzer", *J. Math. Biology*, Vol. 15, pp. 267-273, 1982.

[67] E. Oja, "Neural Networks, Principal Components, and Subspaces", *Int. J. Neural Systems*, Vol. 1, pp. 61-68, 1989.

[68] T. D. Sanger, "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network", *Neural Networks* , Vol. 12, 459-473, 1989.

[69] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

[70] J. P. Basart, M. S. Chackalackal, R. C. Gonzalez, "Introduction to Gray-Scale Morphology", in *Advances in Image Analysis*, edited by Y. Mahdavieh and R. C. Gonzalez, SPIE, Bellingham, WA, 1992.

[71] H. J. A. M. Heumans, "Mathematical Morphology: A Modern Approach in Image Processing Based on Algebra and Geometry", *SIAM Review*, Vol. 37, PP. 1-36, 1995.

[72] M. A. Mohamed and J. Saniie, "Statistical Evaluation of Sequential Morphological Operations", *IEEE Trans. Signal Processing*, Vol. 43, July 1995.

[73] H. Park, and R. T. Chin, "Decomposition of Arbitrarily Shaped Morphological Structuring Elements", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 17, pp. 2-15, 1995.

[74] A. Morales and R. Acharya, "Statistical Analysis of Morphological Openings", *IEEE Trans. Signal Processing*, Vol. 41, pp. 3052-3056, 1993.

[75] M.-H. Chen and P.-F. Yan, "A Multiscaling Approach Based on Morphological Filtering", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, pp. 694-700, 1989.

[76] A. Ye and D. Casasent, "Morphological and Wavelet Transforms for Object Detection and Image Processing", *Applied Optics*, Vol. 33, pp. 8226-8239, 1994.

[77] B. Wang, J. P. Basart, and J. C. Moulder, "The Eddy Current Inverse Problem of a Small Spherical Flaw", in proceedings of the 4th *Annual Midwest Electro-Technology Conference*, pp. 104-107, 1995.

[78] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1992.

[79] B. Wang, J. P. Basart, and J. C. Moulder, "Eddy Current Image Restoration Using Wiener Filtering and Maximum Entropy Method", *in Proc. of the Fifth Annual Midwest Electro-Technology Conference*, Ames, Iowa, April, 1996.

[80] S. F. Gull and J. Skilling, "Maximum Entropy Method in Image Processing", *IEE Proc.* Vol. 131, Pt. F, pp. 646-659, 1984.

[81] S. F. Gull and T. J. Newton, "Maximum Entropy Tomography", *Applied Optics*, Vol. 25, pp. 156-160, 1986.

[82] N. L. Bonavito, J. E. Dorband, and T. Busse, "Maximum Entropy Restoration of Blurred and Oversaturated Hubble Space Telescope Imagery", *Applied Optics,* Vol. 32 pp. 5768-5774, 1993.

[83] A. C. Kokaram, N. Persad, J. Lasenby, W. J. Fitzgerald, A. Mckinnon, and M. Welland, "Restoration of Images from the Scanning-Tunneling Microscope", *Applied Optics,* Vol. 34, pp. 5121-5132, 1995.

[84] E. D. Meinel, "Maximum-entropy Image Restoration: Lagrange and Recursive Techniques", *J. Opt. Soc. Am. A*, Vol. 5, pp. 25-29, 1988.

[85] E. H. L. Aarts, *Simulated Annealing and Boltzmann Machine: a Stochastic Approach to Combinatorial Optimization and Neural Computing*, Wiley, New York, 1989.

[86] V. Granville, M. Krivanek, and J. P. Rasson, "Simulated Annealing: A Proof of Convergence", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 16, pp. 652-656, 1994.

[87] P. F. Jones, B. Lim, and G. J. M. Aitken, "Area-Adaptive Simulated Annealing for Image Reconstruction and Restoration", *Applied Optics,* Vol. 33, pp. 3732-3739, 1994.

[88] K. V. Prasad, R. J. Mammone, and J. Yogeshwar, "Three-dimensional Image Restoration Using Constrained Optimization Techniques", *Optical Engineering*, Vol. 29, pp. 279-288, 1990.

[89] R. K. Chunduru, M. K. Sen, and P. L. Stoffa, "2-D Resistivity Inversion Using Spline Parameterization and Simulated Annealing", *Geophysics*, Vol. 61, pp. 151-161, 1996.

[90] H. Deng, "Synthesis of Binary Sequences with Good Autocorrelation and Crosscorrelation Properties by Simulated Annealing", *IEEE Trans. Aerospace Electronic Systems*, Vol. 32, pp. 98-107, 1996.

[91] L. T. Wille, "New Binary Covering Codes Obtained by Simulated Annealing", *IEEE Trans. Information Theory,* Vol. 42, pp. 300-302, 1996.

[92] V. Murino, A. Trucco, C. S. Regazzoni, "Synthesis of Unequally Spaced Arrays by Simulated Annealing", *IEEE Trans. Signal Processing.* Vol. 44, pp. 119-123, 1996.

[93] J. Radecki, J. Konrad, E. Dubois, "Design of Multidimensional Finite-wordlength FIR and IIR Filters by Simulated Annealing", *IEEE Trans. Circuits and Systems, Part-II, Analog and Digital Signal Processing,* Vol. 42, pp. 424-431, 1995.

[94] A. DeGloria, P. Faraboschi, M. Olivieri, "Block Placement with a Boltzmann Machine", *IEEE Trans. Computer Aided Design of Integrated Circuits and Systems,* Vol. 13, pp. 694-701, 1994.

[95] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, pp. 721-741, 1984.

[96] F. Romeo and Sangiovanni-Vincentelli, "Probabilistic Hill Climbing Algorithms: Properties and Applications", *Proc. Chapel Hill Conf. VLSI*, Chapel Hill, pp.393-417, 1985.

[97] B. Hajek, "Cooling Schedules for Optimal Annealing", *Mathematics of Operational Research*, Vol. 13, pp.311-329, 1988.

[98] S.-Y. Lee and K. G. Lee, "Synchronous and Asynchronous Parallel Simulated Annealing with Multiple Markov Chains", IEEE Trans. Parallel and Distributed Systems, Vol. 7, pp. 993-1008, 1996.

[99] M. Srinivas and L. M. Patnaik, "Genetic Algorithms: a Survey", *Computer,* Vol. 27, pp. 17-26, 1994.

[100] K. Gallagher and M. Sambridge, "Genetic Algorithms: a Powerful Tool for Large-Scale Nonlinear Optimization Problems", *Computer & Geoscience*, Vol. 20, pp. 1229-1236, 1994.

[101] J. H. Holland, *Adaptation in Natural and Artificial Systems*, Univ. of Michigan Press, Ann Arbor, Michigan, 1975.

[102] S. A. Billings and G. L. Zheng, "Radial Basis Function Network Configuration Using Genetic Algorithms", *Neural Networks*, Vol. 8, pp. 877-890, 1995.

[103]D. Reynolds, J. Gomatam, "Stochastic Modeling of Genetic Algorithms", *Artificial Intelligence*, Vol. 82, pp. 303-330, 1996.

[104]D. Whitley, S. Dominic, R. Das, and C. W. Anderson, "Genetic Reinforcement Learning for Neurocontrol Problems", *Machine Learning*, Vol. 13, pp. 259-284, 1993.

[105]C. C. Chiu, P. T. Liu, "Image Reconstruction of a Perfectly Conducting Cylinder by the Genetic Algorithm", *IEE Proc. Microwaves, Antennas and Propagat.*, Vol. 143, pp. 249-253, 1996.

[106]S. C. Kramer, R. C. Martin, "Direct Optimization of Gain Scheduled Controllers Via Genetic Algorithms", *J. Guidance, Control, and Dynamics*, Vol. 19, pp. 636-642, 1996.

[107]M. Kishimoto, K. Sakasai, K. Ara, "Solution of Electromagnetic Inverse Problem Using Combinational Method of Hopfield Neural Network and Genetic Algorithm", *J. Appl. Phys.*, Vol. 79, pp. 1-7, 1996.

[108]D. J. Xu, M. L. Daley, "Design of Optimal Digital Filter Using a Parallel Genetic Algorithm", *IEEE Trans. Circuits and Systems, Part II, Analog and Digital Signal Processing*, Vol. 42, pp. 673-675, 1995.

[109]B. Bhanu, S. Lee, S. Das, "Adaptive Image Segmentation Using Genetic and Hybrid Methods", *IEEE Trans. Aerospace and Electronic Systems*, Vol. 31, pp. 1268-1291, 1995.

BIOGRAPHICAL SKETCH

Bing Wang was born on September 6, 1971 in Neijiang, Sichuan province, China. He received the Bachelor of Science in Electrical Engineering from University of Science and Technology of China in 1991 and the Master of Science in Electronic Engineering from University of Electronic Science and Technology of China in 1993. Before he came to Iowa State University, he worked as a DSP Systems Engineer at the Institute of Electronic System Engineering Design, Chengdu, China. He has served as a Research Assistant in the Department of Electronic Engineering at University of Electronic Science and Technology of China and in the Department of Electrical and Computer Engineering at Iowa State University.

Bing Wang is a member of Tau Beta Pi, Phi Kappa Phi, and IEEE. He received the best paper award of the Fifth Annual Midwest Electro-Technology Conference, Ames, Iowa, April, 1996.