Defect Detection in Correlated Noise

Aleksandar Dogandži and Nawanat EuaAnant

Citation: AIP Conference Proceedings **700**, 628 (2004); doi: 10.1063/1.1711680
View online: http://dx.doi.org/10.1063/1.1711680
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/700?ver=pdfcov
Published by the AIP Publishing

DEFECT DETECTION IN CORRELATED NOISE

Aleksandar Dogandžić and Nawanat Eua-Anant

Iowa State University, Center for Nondestructive Evaluation,
1915 Scholl Road, Ames, IA 50011, USA

ABSTRACT. We present methods for detecting NDE defect signals in correlated noise having unknown covariance. The proposed detectors are derived using the statistical theory of generalized likelihood ratio (GLR) tests and multivariate analysis of variance (MANOVA). We consider both real and complex data models. To allow accurate estimation of the noise covariance, we incorporate secondary data containing only noise into detector design. Probability distributions of the GLR test statistics are derived under the null hypothesis, i.e. assuming that the signal is absent, and used for detector design. We apply the proposed methods to simulated and experimental data and demonstrate their superior performance compared with the detectors that neglect noise correlation.

INTRODUCTION

In nondestructive evaluation (NDE) applications, correlated noise is typically caused by

- backscattered grain noise in ultrasonic NDE systems [1] and
- random liftoff variations in eddy-current systems [2].

Accounting for noise correlation in material inspection can significantly improve defect detection performance. We propose methods for detecting defects in two-dimensional images with correlated noise. The noise is assumed to be correlated between rows of data matrices having unknown covariance. The proposed detectors are derived using the statistical theory of generalized likelihood ratio (GLR) tests and multivariate analysis of variance (MANOVA) (see [3] for a tutorial presentation of MANOVA and [4, Ch. 6.4.2] for the definition of the GLR test). We consider real and complex data models and single- and multiple-trial measurement scenarios. For each data matrix under test, we assume that a noise-only matrix is available and utilize both the data and noise-only matrices to estimate the noise covariance. To decide if a defect is present, the GLR test statistic is compared with a threshold. We derive exact and approximate probability distributions of the GLR test statistics under the null hypothesis (signal absent) and use them to find the threshold that guarantees a specified probability of false alarm.

The paper is organized as follows. We first introduce signal and noise models under single- and multiple-trial scenarios. We then present the GLR detectors for real and complex data models. Finally, we apply the proposed methods to simulated eddy-current and experimental ultrasonic data, compare them with energy detectors that do not account for noise correlation, and conclude the paper by outlining suggestions for future work.

CP700, *Review of Quantitative Nondestructive Evaluation Vol. 23,* ed. by D. O. Thompson and D. E. Chimenti

SIGNAL AND NOISE MODELS

We first present signal and noise models for a single experiment (trial) and then extend them to the multiple-trial scenario.

Single trial: Consider the problem of detecting the presence of a defect signal in an $m \times d$ data matrix under test $\boldsymbol{Y}_{\mathrm{T}}$. We also assume that a noise-only data matrix $\boldsymbol{Z}$ of size $m \times (N-d)$ is available. If we do not have *any* additional information about the nature of the defect signal, we can choose a *nonparametric* model for the signal mean:

$$\mathrm{E}\left[\boldsymbol{Y}_{\mathrm{T}}\right] = \boldsymbol{X}, \tag{1}$$

where $\boldsymbol{X}$ is a matrix of unknown parameters and E[·] denotes expectation. For real measurements, we further model the columns of $\boldsymbol{Y}_{\mathrm{T}}$ as independent Gaussian vectors with an *unknown* $m \times m$ positive definite covariance matrix $\boldsymbol{\Sigma}$. The columns of the noise-only matrix $\boldsymbol{Z}$ are assumed to be independent zero-mean Gaussian vectors with covariance $\boldsymbol{\Sigma}$. Similarly, for complex measurements, we assume that (1) holds, the columns of $\boldsymbol{Y}_{\mathrm{T}}$ are independent circularly symmetric complex Gaussian vectors with an unknown positive definite covariance $\boldsymbol{\Sigma}$, and the columns of $\boldsymbol{Z}$ are independent circularly symmetric zero-mean complex Gaussian vectors with covariance $\boldsymbol{\Sigma}$.

Multiple Trials: In some NDE applications, the experiment in which the measurements are collected is repeated K times to improve the signal-to-noise ratio (SNR). Denote the data matrix under test in the kth trial by $\boldsymbol{Y}_{\mathrm{T},k}$ and the corresponding noise-only matrix by $\boldsymbol{Z}_k$, where $k = 1, 2, \ldots, K$. We assume that the defect signal is the same in all trials (independent of k), i.e.

$$\mathrm{E}\left[\boldsymbol{Y}_{\mathrm{T},k}\right] = \boldsymbol{X}, \quad k = 1, 2, \ldots, K \tag{2}$$

and that the noise is independent between trials and has the same covariance $\boldsymbol{\Sigma}$ (independent of k) in each trial. We also assume that

$$NK \geq m + d,$$

which is needed to ensure that we can estimate $\boldsymbol{\Sigma}$. This condition follows from [5, App. C] and [6, App. B], see also [3, eq. (4)].

GENERALIZED LIKELIHOOD RATIO TESTS

We develop GLR tests for detecting defects based on the above measurement model under both real and complex data scenarios. The GLR tests are useful if signal and noise parameters are unknown, see [4].

Real data: Assuming real measurements, we derive the GLR test for detecting the presence of a defect signal, i.e. testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{X} = \boldsymbol{0}$ (no defect present) versus the alternative $\mathcal{H}_1 : \boldsymbol{X} \neq \boldsymbol{0}$ (defect present). To detect the defect signal, we utilize both the data matrices under test $\boldsymbol{Y}_{\mathrm{T},k}$ and noise matrices $\boldsymbol{Z}_k$. The GLR test computes the ratio of likelihood functions under the two hypotheses, with unknown parameters ($\boldsymbol{X}$ and $\boldsymbol{\Sigma}$ under $\mathcal{H}_0$ and $\boldsymbol{\Sigma}$ under $\mathcal{H}_1$) replaced by their maximum likelihood (ML) estimates. First, define the sufficient statistics for estimating $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$:

$$\overline{\boldsymbol{Y}}_{\mathrm{T}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{Y}_{\mathrm{T},k}, \tag{3a}$$

$$\widehat{\boldsymbol{R}} = \frac{1}{NK}\sum_{k=1}^{K} (\boldsymbol{Y}_{\mathrm{T},k}\boldsymbol{Y}_{\mathrm{T},k}^{T} + \boldsymbol{Z}_k\boldsymbol{Z}_k^{T}), \tag{3b}$$

where "^{T}" denotes a transpose. The GLR test compares

$$\mathrm{GLR} = \frac{|\widehat{\boldsymbol{R}}|}{|\widehat{\boldsymbol{R}} - (1/N)\cdot \overline{\boldsymbol{Y}}_{\mathrm{T}} \overline{\boldsymbol{Y}}_{\mathrm{T}}^{\,T}|} = \frac{1}{|\boldsymbol{I}_d - (1/N)\cdot \overline{\boldsymbol{Y}}_{\mathrm{T}}^{\,T} \widehat{\boldsymbol{R}}^{-1} \overline{\boldsymbol{Y}}_{\mathrm{T}}|} \tag{4}$$

with a threshold, where the presence of a defect is declared if GLR is greater than a threshold. Here $|\cdot|$ denotes the determinant. The above test can be derived using the results of [5] and [6], where estimation and detection algorithms were developed for a more general measurement model and applied to the analysis of evoked responses using electroencephalography/magnetoencephalography (EEG/MEG) arrays. Interestingly, in the scalar case and if the zero-mean data is not available (i.e. $m = d = N = 1$) and after the monotone transformation $\sqrt{\mathrm{GLR} - 1}$, the GLR expression (4) reduces to the familiar t-test: $\overline{y}/\sqrt{s^2}$, where $\overline{y} = \sum_{k=1}^{K} y_k/K$ and $s^2 = \sum_{k=1}^{K}(y_k - \overline{y})^2/K$.

Complex data: For complex measurements, the sufficient and GLR test statistics follow by replacing "^{T}" in (3) and (4) with "^{H}", respectively, where "^{H}" denotes the Hermitian (conjugate) transpose.

GLR Distribution Under $\mathcal{H}_0$

Real data: For real measurements and under $\mathcal{H}_0$, the probability distribution of $1/\mathrm{GLR}$ is (see e.g. [7])

$$\frac{1}{\mathrm{GLR}} \sim \lambda(m, NK - d, d).$$

where λ denotes the Wilks' lambda distribution. The above Wilks' lambda distribution is the distribution of the product of m independent beta random variables with parameters $(\frac{1}{2}(NK - d - i + 1), \frac{1}{2}d)$, $i = 1, 2, \ldots, m$. Since Wilks' lambda distribution does not depend on the unknown parameters ($\boldsymbol{\Sigma}$ in this case), we can compute a threshold τ that maintains a constant probability of false alarm. Such a detector is referred to as a *constant false-alarm rate* (CFAR) detector, see e.g. [4]. For large $NK - d$ (i.e. $NK - d \gg \max\{m, d\}$), the following approximation can be used to compute τ for a specified false-alarm probability (see e.g. [7, Cor. 4.2.1]):

$$P_{\mathrm{FA}} = P\{[NK - \tfrac{1}{2}(d + m + 1)] \cdot \ln \mathrm{GLR} \geq \tau\} \approx P\{\chi^2_{md} \geq \tau\}, \tag{5}$$

where χ^2_{md} denotes a χ^2 random variable with md degrees of freedom.

Complex data: For complex measurements and under $\mathcal{H}_0$, $1/\mathrm{GLR}$ follows the complex Wilks' lambda distribution, see e.g. [8]. As in the real case, the resulting detector is CFAR. For large $NK - d$, the following approximation can be used to compute the threshold τ for a specified false-alarm probability:

$$P_{\mathrm{FA}} = P\{(2NK - d - m - 1) \cdot \ln \mathrm{GLR} \geq \tau\} \approx P\{\chi^2_{2md} \geq \tau\}. \tag{6}$$

Mean-Data Energy Detector

In all numerical examples, we compare the proposed method with the CFAR energy detector (ED) for the mean data $\overline{\boldsymbol{Y}}_{\mathrm{T}}$ and $\overline{\boldsymbol{Z}} = \sum_{k=1}^{K} \boldsymbol{Z}_k/K$. For real measurements, the ED compares

$$\mathrm{ED} = \frac{N - d}{d} \cdot \frac{\mathrm{tr}(\overline{\boldsymbol{Y}}_{\mathrm{T}} \overline{\boldsymbol{Y}}_{\mathrm{T}}^{\mathrm{T}})}{\mathrm{tr}(\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^{\mathrm{T}})} \tag{7}$$

with a threshold τ_{ED}, where the presence of a defect is declared if $\mathrm{ED} > \tau_{\mathrm{ED}}$. The above test statistic has an SNR interpretation. The numerator in (7) is simply the sum of squared

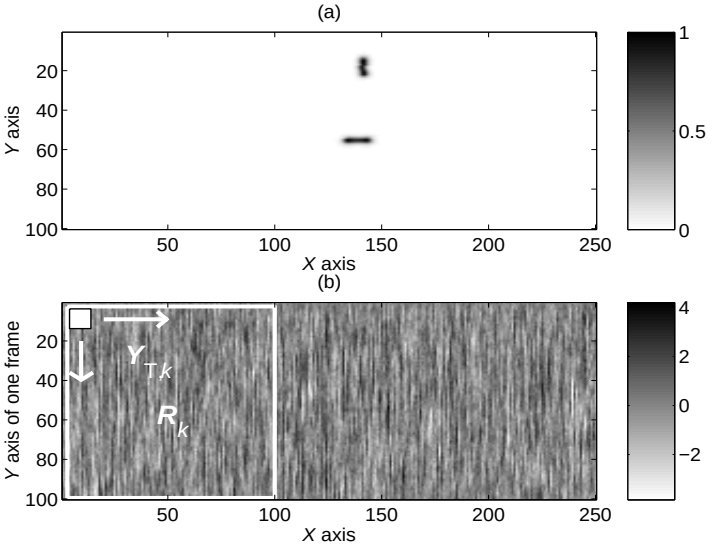

FIGURE 1. (a) Magnitude plot of low-noise eddy-current measurements with peak value normalized to one and (b) a sweeping window $\boldsymbol{Y}_{\mathrm{T},k}$ and a noise-only region $\boldsymbol{R}_k$ shown over noisy measurements for one trial.

magnitudes of the mean data $\overline{\boldsymbol{Y}}_{\mathrm{T}}$, which is an estimate of the overall power in the window under test. Similarly, the denominator in (7) is the sum of squared magnitudes of the (mean) noise-only data $\overline{\boldsymbol{Z}}$, which is an estimate of the noise power. Note that the ED does not account for noise correlation. Under $\mathcal{H}_0$ and if the noise is white, ED is distributed as

$$\mathrm{ED} \sim F(md, m(N-d)), \tag{8}$$

where $F(p,q)$ denotes the F distribution with parameters p and q. For complex measurements, the mean-data energy detector follows by replacing "^{T}" in (7) with "^{H}"; then the distribution of ED under $\mathcal{H}_0$ becomes $\mathrm{ED} \sim F(2md, 2m(N-d))$.

EXAMPLES

Simulated eddy-current data: We consider a simulation example with $K = 10$ trials. Figure 1(a) shows a magnitude plot of low-noise experimental eddy-current impedance measurements in a sample containing two realistic flaws, where each pixel corresponds to a measurement location. The data was collected by scanning the testpiece surface columnwise (parallel to the *y* axis). To model liftoff variations, we added Gaussian noise, correlated along *y* direction (i.e. between rows) and uncorrelated along *x* direction (i.e. independent columns). Matrices $\boldsymbol{Z}_k$, where $k = 1, 2, \ldots, 10$ were generated using noise-only regions $\boldsymbol{R}_k$. Windows $\boldsymbol{Y}_{\mathrm{T},k}$ of size $m \times d = 10 \times 10$ were swept across the noisy images, as shown in Figure 1(b). Figure 2 shows the impedance magnitudes averaged over 10 trials. For each location of the window, we computed the (logarithms of)

- the proposed GLR test statistic in (4) and
- the mean-data energy detector for white noise in (7),

see Figure 3. The results of the GLR and ED detectors are shown in Figure 4. For the probability of false alarm $P_{\mathrm{FA}} = 1\%$, the GLR threshold was computed using (5) and the

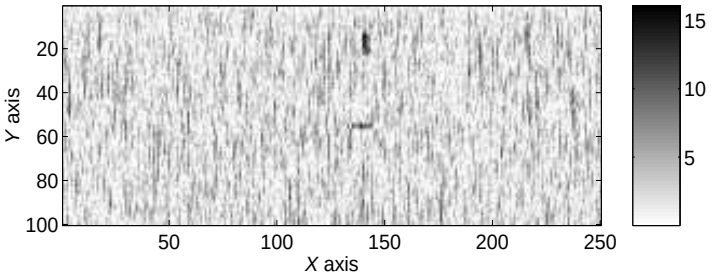

FIGURE 2. Magnitude plot of average data.

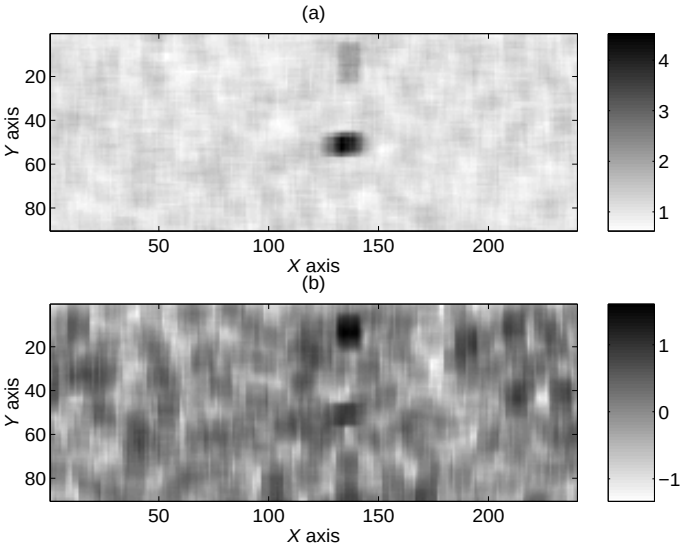

FIGURE 3. Logarithms of (a) the proposed GLR test statistic and (b) the classical ED for multiple trials.

ED threshold was computed by utilizing (8). Black pixels correspond to the test values larger than the threshold. Clearly, the proposed GLR detector, which accounts for noise correlation, outperforms the mean-data energy detector, which breaks down in this scenario.

Experimental ultrasonic data: We applied the GLR and ED tests to ultrasonic C-scan data from an inspection of a cylindrical Ti 6-4 billet, shown in Figure 5. The boxes in Figure 5(a) indicate true defect locations. The data was collected in a single experiment by moving a probe along the axial direction and scanning the billet along the circumferential direction at each axial position. The vertical coordinate is proportional to rotation angle and the horizontal coordinate to axial position. Windows $\boldsymbol{Y}_{\mathrm{T}}$ and $\boldsymbol{W}$ of dimensions $m \times d = 5 \times 5$ and $m \times (N-d) = 5 \times 35$ were swept across the noisy image, as shown in Figure 5(a). We selected the window $\boldsymbol{W}$ from the data that was previously tested and declared to contain only noise. Furthermore, we assume that the noise covariance is a *centrosymmetric* matrix (see [9, Ch. 2.3.7]), i.e. the noise properties are the same for increasing and decreasing y coordinates. Hence, to exploit the stationarity and improve estimation of the noise covariance $\boldsymbol{\Sigma}$, we generate a noise-only matrix $\boldsymbol{Z}$ using both $\boldsymbol{W}$ and a vertically flipped version of $\boldsymbol{W}$.

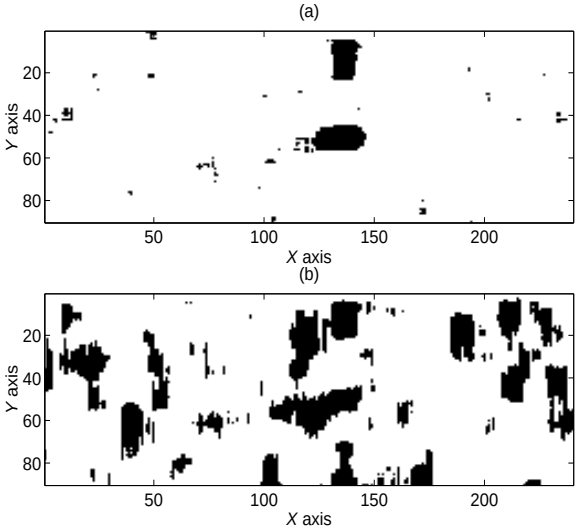

FIGURE 4. GLR detector for multiple trials for $P_{\mathrm{FA}} = 1\%$.

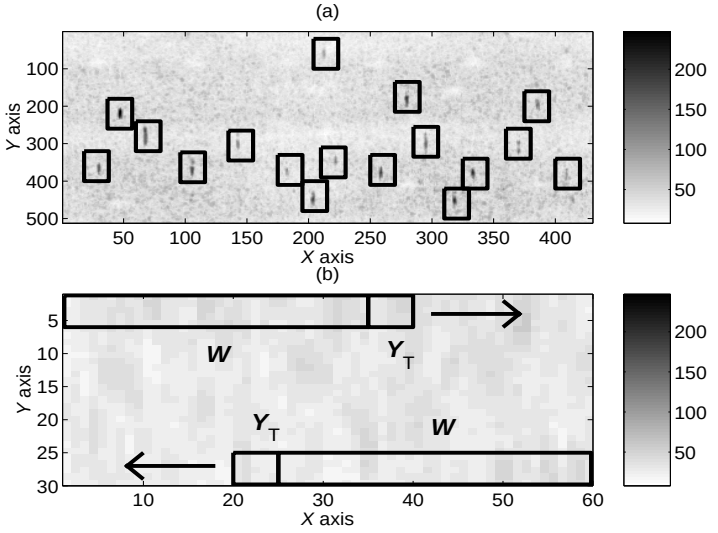

FIGURE 5. (a) Magnitude plot of ultrasonic C-scan data with 17 defects and (b) sweeping window $\boldsymbol{Y}_{\mathrm{T}}$ and a noise-only region $\boldsymbol{W}$ shown over ultrasonic C-scan data.

For each location of the window, we computed the (logarithms of) the proposed GLR test statistic and the energy detector. The windows were swept from left to right and backward as shown in Figure 5(b). For the same $\boldsymbol{Y}_{\mathrm{T}}$, denote the GLRs from the first and the second sweeps by GLR_1 and GLR_2, respectively. Similarly, denote the corresponding EDs by ED_1 and ED_2. Figure 6 shows $[\ln(\mathrm{GLR}_1) + \ln(\mathrm{GLR}_2)]/2$ and $[\ln(\mathrm{ED}_1) + \ln(\mathrm{ED}_2)]/2$, respectively. The presence of a defect is declared if both GLR_1 and GLR_2 (or ED_1 and ED_2) are greater than a threshold. The detection results are shown in Figure 7 for the GLR test and energy detectors, where the threshold was chosen to guarantee the probability of false alarm of 1%. As before, black pixels correspond to the GLR (or ED) values larger than the threshold. Clearly, the proposed GLR detector outperforms the energy detector.

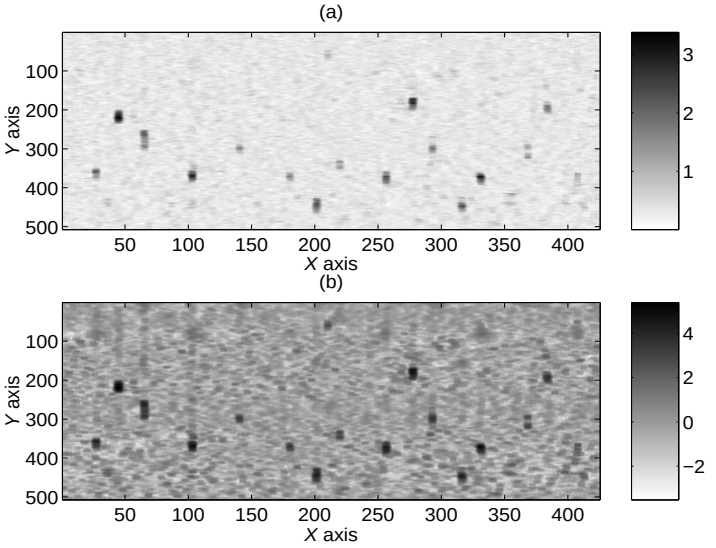

FIGURE 6. The average of logarithms of (a) GLR_1 and GLR_2 and (b) ED_1 and ED_2.

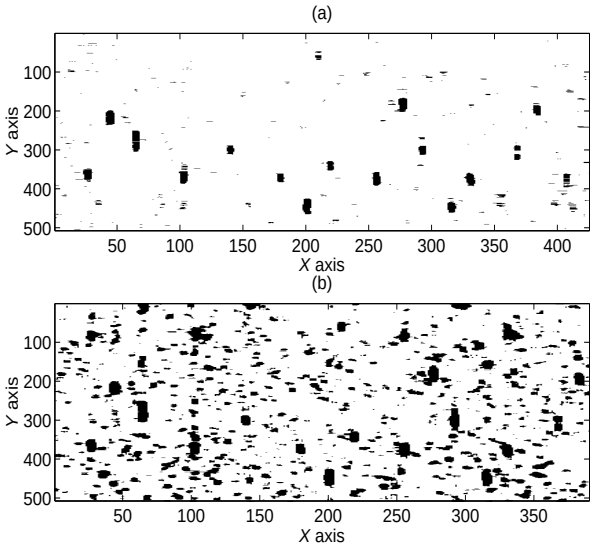

FIGURE 7. (a) GLR detector and (b) energy detector for $P_{\mathrm{FA}} = 1\%$

CONCLUSIONS

We developed generalized likelihood ratio tests for NDE defect detection in correlated noise having unknown covariance. Detectors for real and complex data models were derived and probability distributions of the detectors were computed under the noise-only scenario and used to determine detection thresholds for specified false-alarm probabilities. The proposed detectors were applied to simulated eddy-current and experimental ultrasonic data and compared with the constant false-alarm rate energy detector. Further work will include

- developing the GLR detector for a more general signal-mean model and
- accounting for noise correlation between columns of the data matrices.

ACKNOWLEDGMENT

This work was supported by the NSF Industry-University Cooperative Research Program, Center for Nondestructive Evaluation (CNDE), Iowa State University. The authors are grateful to Drs. R.B. Thompson and J.N. Gray from CNDE and Dr. W. Hassan from Honeywell International Inc. for providing the experimental data.

REFERENCES

1. Thompson, R.B. and Gray, T.A., *Phil. Trans. R. Soc. Lond.* **320**, 329–340 (1986).
2. Auld, B.A. and Moulder, J.C., *J. Nondestructive Eval.* **18**, 3–36 (1999).
3. Dogandžić, A. and Nehorai, A., *IEEE Signal Processing Magazine* **20**, 39–54 (2003).
4. Kay, S.M., *Fundamentals of Statistical Signal Processing: Detection Theory,* Englewood Cliffs, NJ: Prentice Hall, 1998, pt. II.
5. Dogandžić, A. and Nehorai, A., *IEEE Trans. Signal Processing* **48**, 13–25 (2000).
6. Dogandžić, A. and Nehorai, A., "EEG/MEG spatio-temporal dipole source estimation and array design," in *High-Resolution and Robust Signal Processing,* A.B. Gershman and Y. Hua, Eds., New York: Marcel Dekker, 2003, pp. 393–442.
7. Srivastava, M.S. and Carter, E.M., *An Introduction to Applied Multivariate Statistics,* New York, North-Holland, 1983.
8. Kelly, E.J. and Forsythe, K.M., "Adaptive detection and parameter estimation for multidimensional signal models", Lincoln Lab., Mass. Inst. Technol., Lexington, MA, Tech. Rep. 848, 1989.
9. M.H. Hayes, *Statistical Digital Signal Processing and Modeling*, New York: Wiley, 1996.