On the Rate of Growth of Condition Numbers for Convolution Matrices

FAUSTO MILINAZZO, CEDRIC ZALA, AND IAN BARRODALE

Abstract—**When analyzing linear systems of equations, the most important indicator of potential instability is the condition number of the matrix. For a convolution matrix W formed from a series w (where $W_{ij} = w_{i-j+1}$, $1 \le i - j + 1 \le k$, $W_{ij} = 0$ otherwise), this condition number defines the stability of the deconvolution process. For the larger convolution matrices commonly encountered in practice, direct computation of the condition number (e.g., by singular value decomposition) would be extremely time consuming. However, for convolution matrices, an upper bound for the condition number is defined by the ratio of the maximum to the minimum values of the amplitude spectrum of w. This bound is infinite for any series w with a zero value in its amplitude spectrum; although for certain such series, the actual condition number for W may in fact be relatively small. In this paper we give a new simple derivation of the upper bound and present a means of defining the rate of growth of the condition number of W for a band-limited series by means of the higher order derivatives of the amplitude spectrum of w at its zeros. The rate of growth is shown to be proportional to m^p, where m is the column dimension of W and p is the order of the zero of the amplitude spectrum.**

INTRODUCTION

Convolution is a central process in scientific modeling and analysis, and so the inverse problem of deconvolution is frequently encountered in practice. In the discrete case, the convolution of a series w of length k elements with a series s of length m elements to form a series t of length n $(= k + m - 1)$ elements may be expressed in matrix notation as

$$Ws = t,$$

where the elements of the $n \times m$ convolution matrix W are defined by $W_{ij} = w_{i-j+1}$, for $1 \le i - j + 1 \le k$ and $W_{ij} = 0$ otherwise; hence,

$$W = \underbrace{\begin{bmatrix} w_1 & & & & & 0 \\ w_2 & w_1 & & & & \\ w_3 & w_2 & w_1 & & & \\ \vdots & w_3 & w_2 & & & \\ w_k & \vdots & w_3 & w_1 & & \\ & w_k & \vdots & \vdots\ w_2 & w_1 & \\ & & w_k & w_3 & w_2 & w_1 \\ & & & \vdots & w_3 & w_2 \\ & & & w_k & \vdots & w_3 \\ & & & & w_k & \vdots \\ & 0 & & & & w_k \end{bmatrix}}_{m} \Bigg\updownarrow n$$

A general problem in many processing applications is to deconvolve the series t, given the series w, to yield an estimate s' of the series s. In this problem, the numerical condition of the matrix W is of great importance, as it determines the stability of the estimate s'. The condition number $K(W)$ may be used to define an upper limit on the relative change in the deconvolved series s' which may result from a relative change in the convolved series t, and small perturbations in W as follows:

$$\frac{\|\delta s'\|}{\|s'\|} \le \frac{K(W)}{1 - \|\delta W\|\,\|W\|}\left[\frac{\|\delta t\|}{\|t\|} + \frac{\|\delta W\|}{\|W\|}\right]$$

where $\|\cdot\|$ here denotes the l_2 (Euclidean) norm; $K(W)$ is then also defined using this norm. Thus, for a particular choice of w, it is desirable to obtain an estimate of $K(W)$.

One procedure for computing $K(W)$ is through eigenvector–eigenvalue decomposition (EVD) of the correlation matrix W^TW. Then $K(W) = \sqrt{\lambda_{\max}/\lambda_{\min}}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are, respectively, the largest and smallest eigenvalues of W^TW.

Manuscript received August 27, 1986.
F. Milinazzo is with the Mathematics Department, Royal Roads Military College, F.M.O., Victoria, B.C., Canada V0S 1B0.
C. Zala and I. Barrodale are with the Computer Science Department, University of Victoria, P.O. Box 1700, Victoria, B.C., Canada V8W 2Y2.
IEEE Log Number 8612885.

0096-3518/87/0400-0471$01.00 © 1987 IEEE

Alternatively, $K(W)$ may be obtained through singular value decomposition (SVD) of W to yield $W = USV^T$, where U is an $n \times n$ orthogonal matrix, S is an $n \times m$ diagonal matrix with diagonal elements σ_i termed singular values, and V is an $m \times m$ orthogonal matrix. Then $K(W) = \sigma_{\max}/\sigma_{\min}$, i.e., the ratio of the largest to the smallest singular values. Note that the singular values $\sigma_{\max}$ and $\sigma_{\min}$ are the nonnegative square roots of the eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$ of W^TW, and that $K(W) = \sqrt{K(W^TW)}$.

However, in many applications, W is very large (n, $m \approx 1000$) and almost square so that estimation of $K(W)$ by either EVD or SVD is prohibitive. In such cases, it is useful to define an upper bound on $K(W)$, which is approached as the column dimension m of W becomes large. Ekstrom [1], using the results of Grenader and Szego [2] for Toeplitz matrices, has shown that this upper bound in the l_2 norm is given by the ratio of the maximum to the minimum values of the amplitude spectrum of w. In related work, Cybenko [3] and Koltracht and Lancaster [4] derive other expressions for this bound in the l_∞ norm, in terms of the partial correlation coefficients arising in the Levinson–Durbin algorithm.

In an independent analysis of this problem, we have discovered a new, simple, and self-contained derivation of Ekstrom's result. Our analysis uses a simple relation between W^TW and a diagonal matrix with elements equal to the squared amplitude spectrum of w. An application of the Rayleigh quotient yields the desired result. For completeness, the details of the derivation are given in the Appendix.

This bound explains the well-known and often-observed fact that deconvolution may be an ill-conditioned process when w is band-limited. Another interesting implication of the result is that $K(W)$ does not depend on the phase spectrum of w. Thus, the numerical stability of a deconvolution calculation is independent of the phase of the series w.

When the amplitude spectrum of w is zero at one or more frequencies, however, the bound becomes infinite and cannot be used to estimate $K(W)$. In this case, a procedure for estimating the rate of increase of $K(W)$ with m is desirable; if the growth of $K(W)$ were slow enough for a particular w, deconvolution might be a well-conditioned process even though the upper bound for $K(W)$ was infinite. We demonstrate below that the rate of growth of $K(W)$ can be related to the derivatives of the amplitude spectrum of w at its zero values.

Definition of the Rate of Growth of $K(W)$

We use a result of Parter [5] extended by Kesten [6] on Toeplitz matrices to determine how $K(W)$ increases with m when the amplitude spectrum of w contains one or more zeros.

Consider

$$g(\theta) = \sum_{j=-\infty}^{\infty} c_j e^{ij\theta},$$

a real valued Lebesgue integrable function on $\theta \in [-\pi, \pi]$, and the associated $m \times m$ Toeplitz matrix

$$T_m(g) = \begin{bmatrix} c_0 & c_1 & c_2 & \cdot & \cdot & \cdot \\ c_1 & c_0 & c_1 & \cdot & \cdot & \cdot \\ c_2 & c_1 & c_0 & \cdot & \cdot & \cdot \\ & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & c_0 & c_1 & c_2 \\ \cdot & \cdot & \cdot & c_1 & c_0 & c_1 \\ \cdot & \cdot & \cdot & c_2 & c_1 & c_0 \end{bmatrix}.$$

For the purpose of this study, $T_m(g) = W^TW$ for

$$g(\theta) = W_z(e^{i\theta})\, W_z(e^{-i\theta}) = \left| \sum_{j=1}^{k} w_j e^{i(j-1)\theta} \right|^2,$$

where

$$W_z(z) = w_1 + w_2 z + \cdots + w_k z^{k-1}$$

is the z transform of the wavelet w.

Parter [5] proves the following theorem.

Theorem: Let $g(\theta)$ satisfy the following conditions:

i) $g(\theta)$ is real, continuous, and periodic with period 2π;

ii) $g(\theta_0) = \lambda$, λ is the minimum value of g on $[-\pi, \pi]$, and θ_0 is the only value of θ where this minimum is attained;

iii) $g(\theta)$ has $2p$ continuous derivatives in a neighborhood of $\theta = \theta_0$, with $g^{(k)}(\theta_0) = 0$, $1 \le k < 2p$, $g^{(2p)}(\theta_0) = \beta^2 > 0$; that is, the first $2p - 1$ derivatives of $g(\theta)$ at $\theta = \theta_0$ are zero.

Then $\lambda_{\min}$, the smallest eigenvalue of $T_m(g)$, has the asymptotic expansion

$$\lambda_{\min} = \lambda + \frac{\beta^2}{(2p)!} A m^{-2p} + o(m^{-2p})$$

as m goes to infinity, where A is a constant dependent only on g and p.

Since $W_z(e^{i\theta})$ is symmetric about $\theta = 0$, a minimum taken at $\theta = \theta_0$ will also be taken at $\theta = -\theta_0$. Consequently, the theorem as stated above does not strictly apply. Kesten [6] extends Parter's results, showing the same asymptotic dependence of $\lambda_{\min}$ on m when there are multiple points on $[-\pi, \pi]$ where $g(\theta)$ takes on its minimum value.

From the above expression it can be seen that when the amplitude spectrum has a zero of order p, or equivalently when the z transform of w has a zero on the unit circle (i.e., $\lambda = 0$), an estimate for the growth of $K(W^TW)$ with m is given by m^{2p}; consequently, the growth of $K(W)$ with m will be proportional to m^p.

Discussion

The result above shows that if the amplitude spectrum at a frequency θ_0 is zero, the rate of increase of $K(W)$ with m, the column dimension of W, depends on the higher order derivatives of the amplitude spectrum at θ_0. Thus, for w of appropriate spectral characteristics, the computed value for $K(W)$ may be finite and relatively

TABLE I
DESCRIPTION OF WAVELETS USED IN CONDITION NUMBER STUDIES

Wavelet	Spectral Properties	Number of Elements	Description	Condition Number Bound[a]
A	Broad-band	3	$(1.0, 3.0, 1.0)$	5.00
B	Broad-band	6	$(-0.4, -0.6, 0.2, 1.0, 0.5, 0.2)$	9.53
C	Broad-band except for single zero at $\theta_0 = \pi/3$; Nonzero second derivative at θ_0	3	$(1.0, -1.0, 1.0)$	$0.846 \cdot 10^3$
$D\,(=C*C)$	Broad-band except for single zero at $\theta_0 = \pi/3$; Zero second derivative at θ_0	5	$(1.0, -2.0, 3.0, -2.0, 1.0)$	$0.716 \cdot 10^6$
E	Band-limited	43	$\cos(2\pi(0.1)x)\exp(-0.025x^2),\ -21 \le x \le 21$	$0.265 \cdot 10^9$
F	Band-limited	95	Inverse Fourier transform of amplitude spectrum defined by: $\lvert\hat{w}\rvert = 1.0 - \cos(\pi(0.25 - f)/0.25),\quad 0 \le f \le 0.25$; $0,\quad 0.25 \le f \le 0.5$	$0.252 \cdot 10^9$

All calculations were performed in double precision. Sufficient numbers of elements for wavelets E and F were included to ensure that all values of the wavelet within 10^{-5} of the maximum were retained.

[a]Computed for a 1024-point Fourier transform. The true bounds for wavelets C and D are infinite, but the zero at $\pi/3$ does not correspond here to a Fourier frequency.

low for a particular large W even though the bound for $K(W)$ is infinite. For example, if the second derivative of the amplitude spectrum at θ_0 is nonzero, $K(W)$ may be expected to increase linearly with m; if the first and second (but not the third) derivatives at θ_0 are zero, $K(W)$ may grow as m^2. At the opposite extreme, for band-limited wavelets where a substantial region of the amplitude spectrum is near zero, the higher order derivatives are also near zero, and $K(W)$ will grow as a higher power of m, or even exponentially.

We have performed numerical studies for various model wavelets in which the upper bound for $K(W)$ was calculated from the amplitude spectrum (1024-point Fourier transform) and the rate of increase of $K(W)$ was determined by applying SVD to the matrix W. The wavelets are described in Table I and the results of the experiments are illustrated in Fig. 1.

For the broad-band wavelets A and B, the bounds for $K(W)$ are small and are rapidly approached for W of low column dimension m. The wavelet C has an amplitude spectrum with a single zero at $\theta_0 = \pi/3$, and thus, the true bound for $K(W)$ is infinite. However, the second derivative is nonzero and, as expected from the above analysis, $K(W)$ is comparatively small even for large m, and increases linearly with m. The wavelet D was obtained by convolving the wavelet C with itself, so that its amplitude spectrum is the square of that of the wavelet C; consequently, both the first and second derivatives of the amplitude spectrum at θ_0 are zero. As predicted, $K(W)$ was found to increase as m^2 for this wavelet.

The wavelets E and F were designed to have substantial near-zero regions in their amplitude spectra so that the second and higher order derivatives would also be near zero in these regions. The computed bounds for $K(W)$ for these wavelets were both extremely large ($> 10^8$) but not infinite, because of the necessity of truncating the wavelets in the time domain. Fig. 1 shows that for both

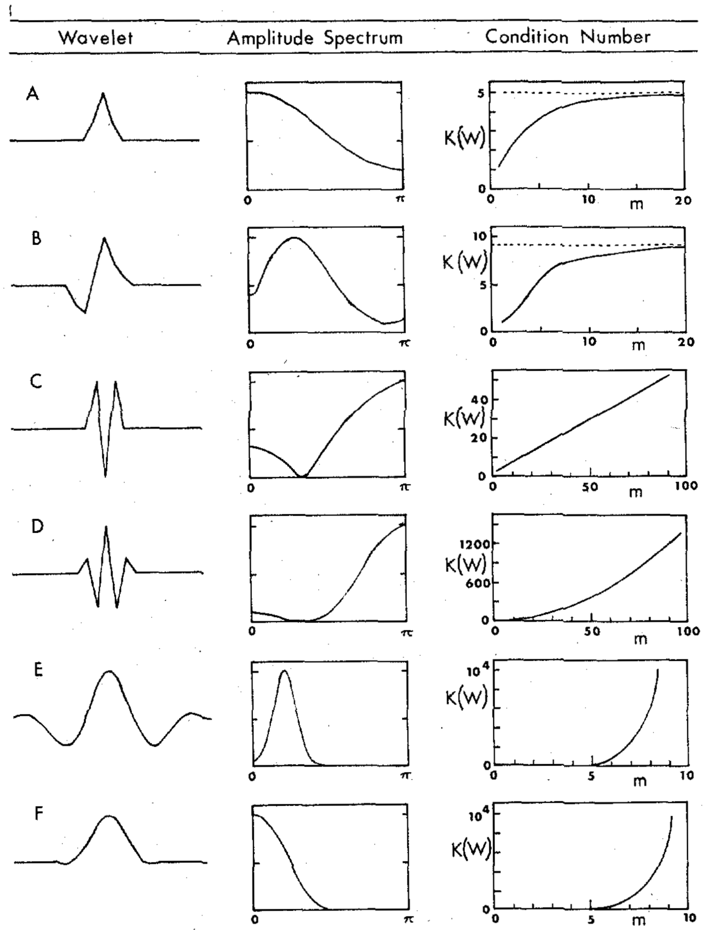


Fig. 1. Wavelets—their amplitude spectra and rate of growth of condition number. The wavelets A to F described in Table I are shown on the left column and their corresponding normalized amplitude spectra appear in the center column. The condition numbers $K(W)$ for the convolution matrices W corresponding to each wavelet were computed by singular value decomposition for W of various column dimension m, and are plotted as a function of m in the column of graphs on the right. The dashed lines in these graphs for wavelets A and B indicate the computed bounds for the condition number. Note that while the scales for the wavelets and the spectra are the same within each column, both the horizontal and vertical scales for the condition number plots may be very different.

wavelets, $K(W)$ increases rapidly with m, reaching a value of more than 10^4 for a 10-column matrix. A log-linear plot of the data confirmed that the rate of increase was exponential, as would be expected for the case where the derivatives of all orders are zero. These matrices are thus exceptionally ill conditioned.

These experiments provide numerical confirmation for the above analysis and illustrate the relationship between the amplitude spectrum of a wavelet and the condition number $K(W)$. While it has been demonstrated that certain series w may be defined for which W is well conditioned even though the bound for $K(W)$ is infinite, for many important applications the amplitude spectrum of w is constant at zero through a substantial region; here $K(W)$ will increase exponentially with m and W may be pathologically ill conditioned even for very small m. Thus, for such deconvolution problems, even those of small size, it is essential to employ algorithms which have been specially designed to address the problem of instability. Such algorithms (for example, see [7] for review) typically incorporate sufficient additional assumptions or constraints so as to result in a solution which is unique and physically realistic.

Appendix
Derivation of an Upper Bound for $K(W)$

Theorem: Let W be an $n \times m$ convolution matrix as previously defined, and let the discrete Fourier transform of w over n elements at frequency v be

$$\hat{w}_v = \sum_{j=1}^{k} w_j \exp\left(-ihv(j-1)\right)$$

for $h = 2\pi/n$ and $v = 0, 1, 2, \cdots, n-1$. Then $K(W)$, the condition number of W, is bounded by

$$K(W) \le \frac{\max_v |\hat{w}_v|}{\min_v |\hat{w}_v|},$$

i.e., by the ratio of the maximum to the minimum value of the amplitude spectrum of w.

Proof: Let F be an $n \times n$ matrix defined as follows:

$$F = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha^{n-1} & \alpha^{2(n-1)} & \cdots & \alpha^{(n-1)^2} \end{bmatrix}$$

where $\alpha = \exp(-ih)$. Then

$$FW = DG,$$

where D is an $n \times n$ matrix and G is an $n \times m$ matrix defined, respectively, as

$$D = \begin{bmatrix} \hat{w}_0 & & & & \mathbf{0} \\ & \hat{w}_1 & & & \\ & & \hat{w}_2 & & \\ & & & \ddots & \\ \mathbf{0} & & & & \hat{w}_{n-1} \end{bmatrix}$$

and

$$G = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{m-1} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(m-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha^{(n-1)} & \alpha^{2(n-1)} & \cdots & \alpha^{(m-1)(n-1)} \end{bmatrix}.$$

Since F is unitary (i.e., $F^*F = I$), it is evident that G, a matrix consisting of the first m columns of F, satisfies the relation

$$G^*G = I \tag{1}$$

where G^* is the conjugate transpose of G. Also, we observe that

$$\begin{aligned} W^TW &= W^TF^*FW \\ &= (FW)^*(FW) \\ &= (DG)^*(DG) \\ &= G^*D^*DG. \end{aligned} \tag{2}$$

At this stage, we note that $K(W) = \sqrt{K(W^TW)}$ is dependent on the amplitude, but not the phase spectrum of W, since the diagonal elements of D^*D involve only the amplitude spectrum.

Equations (1) and (2) may now be used together with the Rayleigh quotient to estimate the eigenvalues of W^TW. Let $\lambda_{\max}$ and $\lambda_{\min}$ be the largest and smallest eigenvalues of W^TW. Then

$$\begin{aligned} \lambda_{\max} &= \max_{x \ne 0} \frac{x^*W^TWx}{x^*x} = \max_{x \ne 0} \frac{x^*G^*D^*DGx}{x^*x} \\ &= \max_{\substack{y = Gx, \\ x \ne 0}} \frac{y^*D^*Dy}{y^*y} \le \max_{y \ne 0} \frac{y^*D^*Dy}{y^*y}. \end{aligned}$$

The inequality arises when we take the maximum over a larger vector space by dropping the constraint $y = Gx$. Since D^*D is simply a diagonal matrix with elements equal to the squared amplitude spectrum $|\hat{w}_v|^2$, we therefore have

$$\lambda_{\max} \le \max_v |\hat{w}_v|^2.$$

In a similar way, we obtain the relation

$$\lambda_{\min} \ge \min_{y \ne 0} \frac{y^*D^*Dy}{y^*y}$$

$$\lambda_{\min} \ge \min_v |\hat{w}_v|^2.$$

Since the condition number of W^TW equals $\lambda_{max}/\lambda_{min}$, we have

$$K(W^TW) = \frac{\lambda_{max}}{\lambda_{min}} \le \frac{\max_v |\hat{w}_v|^2}{\min_v |\hat{w}_v|^2}$$

so that

$$K(W) \le \frac{\max_v |\hat{w}_v|}{\min_v |\hat{w}_v|}. \qquad \text{QED}$$

Using the Rayleigh quotient, it may be seen that while the condition number of the matrix W^TW increases with order m, the bound for $K(W^TW)$ is independent of m.

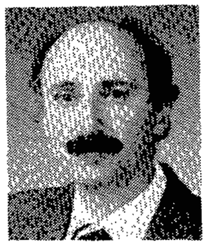
This is also apparent if we note that $\hat{w}_v$ is simply the value of $W_z(z)$, the z transform of the wavelet, at $z = \exp(ihv)$, so that

$$K(W) \le \frac{\max_{|z|=1} |W_z(z)|}{\min_{|z|=1} |W_z(z)|}.$$

The latter quantity is clearly independent of both m and n.

References

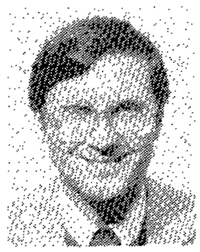
[1] M. P. Ekstrom, "A spectral characterization of the ill-conditioning in numerical deconvolution," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 344–348, Aug. 1973.
[2] U. Grenader and G. Szego, *Toeplitz Forms and Their Applications.* Berkeley, CA: Univ. California Press, 1958.
[3] G. Cybenko, "The numerical stability of the Levinson–Durbin algorithm for Toeplitz systems of equations," *SIAM J. Sci. Stat. Comput.*, vol. 1, pp. 303–319, 1980.
[4] I. Koltracht and P. Lancaster, "Condition numbers of Toeplitz and block Toeplitz matrices," in *Advances in Operator Theory.* Germany: Birkhauser-Verlag, to appear.
[5] S. V. Parter, "On the extreme eigenvalues of truncated Toeplitz matrices," *Bull. Amer. Math. Soc.*, vol. 67, pp. 191–196, 1961.
[6] H. Kesten, "On the extreme eigenvalues of translation kernels and Toeplitz matrices," *J. d'Analyse Math.*, vol. 10, pp. 117–138, 1962.
[7] P. A. Jansson, *Deconvolution: With Applications in Spectroscopy.* Orlando, FL: Academic, 1984.



Fausto Milinazzo received the B.Sc. degree in mathematics and physics in 1970 and the Ph.D. degree in applied mathematics in 1974 from the University of British Columbia, Vancouver, B.C., Canada.

Currently he is an Associate Professor of Mathematics at Royal Roads Military College, Victoria, B.C. His interests include numerical analysis and computation fluid mechanics.

Dr. Milinazzo is a member of the Society for Industrial and Applied Mathematics.



Cedric Zala received the B.Sc. degree in biochemistry from the University of Victoria, B.C., Canada, in 1971 and the Ph.D. degree in biological chemistry from the University of Manchester, Manchester, England, in 1974. During several years of biomedical research at the Lady Davis Institute in Montreal, P.Q., Canada, he became increasingly interested in computing and received a Certificate in computer programming from McGill University.

He has been a Consultant with Barrodale Computing Services since 1981 and has also been involved with research projects through the University of Victoria. His special interest is applications of the L_1 norm in signal processing.

Ian Barrodale received the B.Sc. degree in mathematics from the University of Wales, Bangor, in 1960, the M.A. degree in mathematics from the University of British Columbia, Vancouver, B.C., Canada, in 1965, and the Ph.D. degree in numerical analysis from the University of Liverpool, England, in 1967.

He is a Consultant and a part-time Professor at the University of Victoria, B.C., Canada, where he has taught mathematics and computer science since 1961. He has held visiting positions as a Numerical Analyst at the Mathematics Research Center, Madison, WI, the Atomic Energy Research Establishment, Harwell, England, Liverpool University, and the Defense Research Establishment, Victoria. His research interests are in applied numerical analysis, operations research, and scientific programming. He has been an Editor of the *SIAM Journal on Numerical Analysis* and has served on advisory committees for the National Research Council and the Defense Research Board.