The Performance Surface in Filtered Nonlinear Mean-Square Estimation

Márcio H. Costa, José Carlos M. Bermudez, and Neil J. Bershad

***Abstract*—This brief investigates the properties of the performance surface for the problem of linearly constrained nonlinear mean-square estimation of a random sequence. The problem studied has direct application to the study of active noise control systems when the transducers are driven into nonlinear behavior. A deterministic expression is derived for the mean-square error (MSE) surface as a function of the nonlinearity parameter for Gaussian inputs. It is demonstrated that the surface is unimodal, and expressions are determined for the optimum weight vector and for the minimum MSE.**

***Index Terms*—Adaptive filters, adaptive signal processing, active noise control (ANC), nonlinear systems, estimation theory.**

I. Introduction

Mean-square estimation plays a crucial role in many problems of adaptive control and adaptive signal processing [1], [2]. The optimal design of adaptive estimation systems requires detailed knowledge about the theoretical problem and about the adaptive algorithm performance in solving that problem. Such knowledge is obtained through analysis of the system behavior and derivation of analytical models that can accurately predict the adaptive algorithm behavior when applied to that system.

The study of the adaptive algorithm behavior leads to the prediction of its transient and steady-state behaviors in seeking a stationary point of the performance surface. The steady-state results must then be compared to the stationary points of the performance surface to determine the efficiency of the algorithm. Thus, knowledge of performance surface properties (such as its minimum and the uniqueness of such a minimum) is required to determine the adaptive algorithm's performance, as well as to compare performances of different algorithms. The most employed performance surface is the mean-square error (MSE). In general, the MSE cost function has a second-order dependence on the adaptive filter coefficients, has a unique global minimum, and is mathematically tractable [1].

Several adaptive modeling and control systems present saturation-type nonlinearities in the adaptive filter path [3]–[6]. Such nonlinearities can severely affect the MSE surface properties and the ability of the adaptive algorithm to minimize the MSE. Several works [3], [7], [8] have studied the behavior of adaptive filters in modeling and control applications with such nonlinearities. Active noise control (ANC) is one important example. ANC systems are described in detail in [9]. The frequently used models for the nonlinear adaptive path define the nonlinear mean square estimation problem depicted in

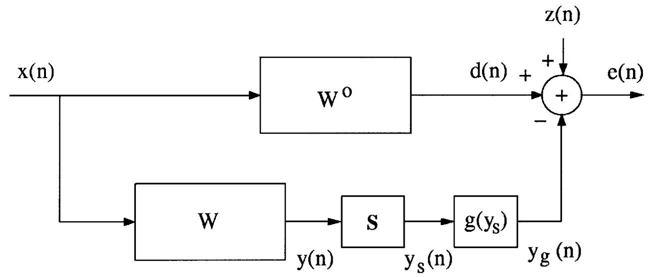

Fig. 1. Constrained mean-square estimation problem.

Fig. 1. The signal $x(n)$ represents the acoustic noise which is directly measurable. This signal $x(n)$ propagates through an unknown linear acoustic medium, which is modeled by W^o. The output of the acoustic medium is generated via the linear model $d(n) = \sum_{k=0}^{N-1} w_k^o x(n-k)$, with order N determined *a priori*,[1] and is obscured by the additive noise $z(n)$. The system W is the linear estimator, whose parameters must be determined to minimize the mean-square value of the error $e(n)$.[2] The system S is a linear filter which models the linear response of the amplifier used to drive the acoustic transducer (usually a speaker). It is known as *the secondary path* in the ANC literature. Its response can be estimated using online or offline techniques [9]. This filtering operation represents a set of constraints on the response of the linear system comprised of a cascade connection of W and S. The nonlinearity $g(\cdot)$ models the amplifier or the transducer saturation effects. Thus, in the system of Fig. 1, $d(n) + z(n)$ is estimated by a nonlinear function of the reference signal $x(n)$ [11, Sec. 7-5].

Although system nonlinearities are quite common, very little has been reported in the literature on their effects on the MSE surface. A recent paper [8] has studied the statistical behavior of the filtered-X least-mean-square (FXLMS) adaptive filter when used to solve the problem in Fig. 1. This analysis determined analytical models for the mean weight and the MSE behaviors. However, the results in [8] alone do not provide all the necessary design information if the MSE performance surface properties are unknown. The knowledge of such properties allows the designer to determine the algorithm behavior for a given degree of nonlinearity, as compared to the optimum. In addition, the MSE surface properties are necessary for a meaningful performance comparison among different adaptive algorithms.

This brief determines the MSE surface properties for systems described by Fig. 1 when the input $x(n)$ is Gaussian. A deterministic expression is derived for the MSE surface as a function of the nonlinearity parameter. The surface is shown to remain unimodal for any degree of nonlinearity. The optimum weight vector and the minimum MSE are determined.

II. Analysis of the MSE Surface

Fig. 1 corresponds to a nonlinear mean-square estimation problem [11, Sec. 7-5]. The sequence $d(n)$ is estimated in the mean square sense by a nonlinear function of the reference signal $x(n)$. In the following analysis, $W^o = [w_0^o\, w_1^o \ldots w_{N-1}^o]^T$, where w_k^o is the kth sample of the impulse response of the unknown system, $W = [w_0\, w_1 \cdots w_{N-1}]^T$ is the weight vector of the linear filter to be optimized, $S = [s_0\, s_1 \cdots s_{M-1}]^T$ contains the samples of the secondary path impulse response, $X(n) = [x(n)\, x(n-1) \cdots x(n-N+1)]^T$ is the observed

[1]The model order can be estimated using different techniques [10, Ch. 16].

[2]In adaptive filtering, W is replaced by a time-varying filter $W(n)$ which is controlled by an adaptive algorithm [9].

Manuscript received January 14, 2002; revised July 13, 2002. This work was supported in part by the Brazilian Ministry of Education (CAPES) under Grant PICDT 0129/97-9 and in part by the Brazilian Ministry of Science and Technology (CNPq) under Grant 352084/92-8. This paper was recommended by Associate Editor O. Au.

M. H. Costa is with the Grupo de Engenharia Biomédica, Escola de Engenharia e Arquitetura, Universidade Católica de Pelotas, Pelotas 96010-000, RS, Brazil (e-mail: m.costa@ieee.org).

J. C. M. Bermudez is with the Department of Electrical Engineering, Federal University of Santa Catarina, Florianópolis 88040-900, SC, Brazil (e-mail: j.bermudez@ieee.org).

N. J. Bershad is with the Department of Electrical and Computer Engineering, University of California at Irvine, Irvine, CA 92697 USA (e-mail: bershad@ece.uci.edu).

Digital Object Identifier 10.1109/TCSI.2003.808845

input data vector, $y(n)$ is the output of W, $y_s(n)$ is the output of S, $g(y_s)$ is the saturation nonlinearity, $y_g(n)$ is the nonlinearity output (canceling signal) and $e(n)$ is the estimation error. The signal $x(n)$ is assumed statistically stationary, zero-mean Gaussian with positive definite correlation matrix. The Gaussian assumption leads to better mathematical tractability. Moreover, Gaussian models are usually appropriate to deal with stochastic processes generated by physical phenomena. The noise $z(n)$ is stationary, white, zero-mean, with variance σ_z^2 and statistically independent of any other signal. Vectors W^o and W are assumed to have the same size to simplify matrix notations. The saturation nonlinearity is modeled by the scaled error function

$$g(y_s) = \int_0^{y_s} e^{-\frac{z^2}{2\sigma^2}}\, dz. \tag{1}$$

Note that $\lim_{\sigma^2 \to \infty} [g(y_s)] = y_s$ and $\lim_{\sigma^2 \to 0} [g(y_s)] = \sigma\sqrt{\pi/2}\,\mathrm{sgn}(y_s)$. Hence, by changing σ^2, $g(y_s)$ can be varied between a linear device and a hard limiter. The effects of very large nonlinearities ($\sigma \to 0$) can be studied by scaling $g(y_s)$ by a constant such as A/σ, where $A \in R_+$. $g(y_s)$ models the saturation-type nonlinearity, which is of great practical interest.

A. MSE Performance Surface

The error signal in Fig. 1 is given by

$$\begin{aligned} e(n) &= d(n) + z(n) - y_g(n) \\ &= W^{o^T} X(n) + z(n) - g\left[\sum_{i=0}^{M-1} s_i W^T X(n-i)\right]. \end{aligned} \tag{2}$$

Squaring $e(n)$ in (2) and taking the expected value yields

$$\begin{aligned} E\{e^2(n)\} = &W^{o^T} E\left\{X(n)X^T(n)\right\} W^o + 2W^{o^T} E\{z(n)X(n)\} \\ &- 2W^{o^T} E\left\{g\left[\sum_{i=0}^{M-1} s_i W^T X(n-i)\right] X(n)\right\} \\ &+ E\{z^2(n)\} - 2E\left\{z(n) g\left[\sum_{i=0}^{M-1} s_i W^T X(n-i)\right]\right\} \\ &+ E\left\{g^2\left[\sum_{i=0}^{M-1} s_i W^T X(n-i)\right]\right\}. \end{aligned} \tag{3}$$

Four expectations in (3) are easily evaluated using the statistical properties of $x(n)$ and $z(n)$: $E\{X(n)X^T(n)\} = R_0$, $E\{z(n)X(n)\} = 0$, $E\{z^2(n)\} = \sigma_z^2$ and $E\{z(n)g[\sum_{i=0}^{M-1} s_i W^T X(n-i)]\} = 0$, where the notation $R_{j-i} = E\{X(n-i)X^T(n-j)\}$ is used. Thus, R_0 is the autocorrelation matrix of $X(n)$.

The third expectation is of the form $E\{g(y_1)Y_2\}$, where y_1 and the components of vector Y_2 are zero-mean Gaussian variates. This expectation can be determined using the Modified Price theorem [12]. Following the same steps in [13, App.A] with $b = 0$, $c = 1/\sigma$ and $\sigma_q = \sigma$, it can be shown that

$$E\left\{g\left[\sum_{i=0}^{M-1} s_i W^T X(n-i)\right] X(n)\right\} = \frac{1}{\sqrt{\frac{1}{\sigma^2} W^T \tilde{R}_{ss} W + 1}} \tilde{R}_s^T W \tag{4}$$

where[3]

$$\tilde{R}_{ss} = \sum_{j=0}^{M-1} \sum_{i=0}^{M-1} s_j s_i R_{j-i} \tag{5}$$

$$\tilde{R}_s = \sum_{i=0}^{M-1} s_i R_{-i}. \tag{6}$$

[3]Note that $\tilde{R}_s^T = \sum_{i=0}^{M-1} s_i R_i$.

The last expectation can be obtained from [14, eq. (40)] for $\alpha_f = b_1(n) = 1$, $H = W$ and $Y(n) = X(n)$:[4]

$$E\left\{g^2\left[\sum_{i=0}^{M-1} s_i W^T X(n-i)\right]\right\} = \sigma^2 \arcsin\left(\frac{W^T \tilde{R}_{ss} W}{W^T \tilde{R}_{ss} W + \sigma^2}\right). \tag{7}$$

Using (4) and (7) in (3) yields an analytical expression for the MSE surface

$$\begin{aligned} \xi(W) = E\{e^2(n)\} = &\sigma_z^2 + W^{o^T} R_0 W^o \\ &- \frac{2}{\sqrt{\frac{1}{\sigma^2} W^T \tilde{R}_{ss} W + 1}} W^{o^T} \tilde{R}_s^T W \\ &+ \sigma^2 \arcsin\left(\frac{W^T \tilde{R}_{ss} W}{W^T \tilde{R}_{ss} W + \sigma^2}\right). \end{aligned} \tag{8}$$

Equation (8) reduces to the MSE expression for the linear case as $\sigma^2 \to \infty$ [15].

B. Stationary Points

In the following, it is assumed that $\tilde{R}_{ss}$ is positive definite, a reasonable assumption for most practical systems [1].[5] Differentiating (8) with respect to W, equating the result to zero and denoting $\tilde{W}$ as the finite values of W that satisfy the resulting equation, it can be easily shown that

$$\tilde{W} = \frac{\left(1 + \frac{1}{\sigma^2} \tilde{W}^T \tilde{R}_{ss} \tilde{W}\right)}{\frac{1}{\sigma^2} W^{o^T} \tilde{R}_s^T \tilde{W} + \frac{\left(\frac{1}{\sigma^2} \tilde{W}^T \tilde{R}_{ss} \tilde{W} + 1\right)^{\frac{1}{2}}}{\left(\frac{2}{\sigma^2} \tilde{W}^T \tilde{R}_{ss} \tilde{W} + 1\right)^{\frac{1}{2}}}} \tilde{R}_{ss}^{-1} \tilde{R}_s W^o. \tag{9}$$

Thus, $\tilde{W} = c\, \tilde{R}_{ss}^{-1} \tilde{R}_s W^o$, where c is a real scalar for any finite W^o and $\tilde{W}$. Note that, contrary to the nonfiltered case [16], $\tilde{W}$ is not a scaled version of W^o. In general, the matrix $\tilde{R}_{ss}^{-1} \tilde{R}_s$ modifies the direction of W^o, as well as its magnitude. Substituting $c\, \tilde{R}_{ss}^{-1} \tilde{R}_s W^o$ for $\tilde{W}$ in (9) and defining

$$\beta^2 = \frac{1}{\sigma^2} W^{o^T} \tilde{R}_s^T \tilde{R}_{ss}^{-1} \tilde{R}_s W^o \tag{10}$$

leads to

$$c\tilde{R}_{ss}^{-1} \tilde{R}_s W^o = \frac{(c^2\beta^2 + 1)}{c\beta^2 + \frac{(c^2\beta^2+1)^{\frac{1}{2}}}{(2c^2\beta^2+1)^{\frac{1}{2}}}} \cdot \tilde{R}_{ss}^{-1} \tilde{R}_s W^o. \tag{11}$$

Equating the scalar multipliers on both sides yields

$$c\sqrt{\frac{c^2\beta^2 + 1}{2c^2\beta^2 + 1}} = 1. \tag{12}$$

Equation (12) shows that c must be positive. Solving (12) for c yields the four solutions

$$c_{1,2,3,4} = \pm\sqrt{1 - \frac{1}{2\beta^2} \pm \sqrt{\frac{1}{4\beta^4} + 1}}. \tag{13}$$

Two of these solutions are complex and one is real negative. Thus, the only solution satisfying $c \in R_+$ is

$$c = \sqrt{1 - \frac{1}{2\beta^2} + \sqrt{\frac{1}{4\beta^4} + 1}} \tag{14}$$

and thus

$$\tilde{W} = \sqrt{1 - \frac{1}{2\beta^2} + \sqrt{\frac{1}{4\beta^4} + 1}} \cdot \tilde{R}_{ss}^{-1} \tilde{R}_s W^o \tag{15}$$

corresponds to the only finite point for which $\partial\xi(W)/\partial W = 0$. Appendix A presents a mathematical proof that the Hessian $\nabla^2[\xi(W)]$ is positive definite at $W = \tilde{W}$. Thus, (15) corresponds to a minimum of $\xi(W)$.

[4]Note that $b = \sigma_y^2 H^T H$ in [14] would be $H^T R_0 H$ for $x(n)$ nonwhite

[5]$\tilde{R}_{ss}$ is the autocorrelation matrix of $x(n)$ filtered by the filter S [15].

Setting $W = \tilde{W}$ in (8) and using (15) yields an expression for the minimum MSE

$$\xi_{\min} = \sigma_z^2 + W^{o^T} R_0 W^o + \left[\frac{1}{\beta^2}\arcsin\left(\frac{c^2\beta^2}{c^2\beta^2+1}\right) - \frac{2c}{\sqrt{c^2\beta^2+1}}\right] \cdot W^{o^T}\tilde{R}_s^T\tilde{R}_{ss}^{-1}\tilde{R}_s W^o \quad (16)$$

where β^2 and c are given by (10) and (14), respectively. Again, as $\beta^2 \to 0$ (toward the linear case), (16) reduces to the linear case optimum solution $\xi_{\substack{\min\\ \text{linear}}} = \sigma_z^2 + W^{o^T} R_0 W^o - W^{o^T}\tilde{R}_s^T\tilde{R}_{ss}^{-1}\tilde{R}_s W^o$, in agreement with [15, eq. (11)].

III. Conclusion

This brief has derived the properties of the performance surface for the problem of linearly constrained nonlinear mean-square estimation of a random Gaussian sequence. The problem studied has direct application to the study of ANC systems when the transducers are driven into a nonlinear behavior. A deterministic expression was derived for the MSE surface as a function of the system's degree of nonlinearity. It was demonstrated that the MSE surface is unimodal, and expressions were determined for the optimum weight vector and for the minimum MSE. The results in this brief contribute to the performance analysis of adaptive algorithms applied to nonlinear filtering problems, such as ANC.

Appendix A

Proof That $\nabla^2\xi(\tilde{W})$ Is Positive Definite

The Hessian of $\xi(W)$ is given by

$$\begin{aligned}\nabla^2\xi(W) =& \frac{2}{\sigma^2\left(\frac{1}{\sigma^2}W^T\tilde{R}_{ss}W+1\right)^{\frac{3}{2}}} \times\left[\tilde{R}_{ss}WW^{o^T}\tilde{R}_s^T + \tilde{R}_s W^o W^T\tilde{R}_{ss}\right]\\ &+\left[\frac{\frac{2}{\sigma^2}W^{o^T}\tilde{R}_s^T W}{\left(\frac{1}{\sigma^2}W^T\tilde{R}_{ss}W+1\right)^{\frac{3}{2}}} + \frac{2}{\left(\frac{2}{\sigma^2}W^T\tilde{R}_{ss}W+1\right)^{\frac{1}{2}}\left(\frac{1}{\sigma^2}W^T\tilde{R}_{ss}W+1\right)}\right]\tilde{R}_{ss}\\ &-\left[\frac{\frac{6}{\sigma^2}W^{o^T}\tilde{R}_s^T W}{\sigma^2\left(\frac{1}{\sigma^2}W^T\tilde{R}_{ss}W+1\right)^{\frac{5}{2}}} + \frac{\frac{12}{\sigma^2}W^T\tilde{R}_{ss}W+8}{\sigma^2\left(\frac{2}{\sigma^2}W^T\tilde{R}_{ss}W+1\right)^{\frac{3}{2}}\left(\frac{1}{\sigma^2}W^T\tilde{R}_{ss}W+1\right)^2}\right]\\ &\times\tilde{R}_{ss}WW^T\tilde{R}_{ss}.\end{aligned} \quad (17)$$

At $\tilde{W} = c\,\tilde{R}_{ss}^{-1}\tilde{R}_s W^o$ and using (10) and (14), (17) becomes

$$\begin{aligned}\nabla^2\xi(\tilde{W}) =& \left[\frac{2c\beta^2(2c^2\beta^2+1)^{\frac{1}{2}} + 2(c^2\beta^2+1)^{\frac{1}{2}}}{(2c^2\beta^2+1)^{\frac{1}{2}}(c^2\beta^2+1)^{\frac{3}{2}}}\right]\tilde{R}_{ss}\frac{1}{\sigma^2}\\ &+\left[\frac{(-2c^3\beta^2+4c)(2c^2\beta^2+1)^{\frac{3}{2}} - (12c^4\beta^2+8c^2)(c^2\beta^2+1)^{\frac{1}{2}}}{(2c^2\beta^2+1)^{\frac{3}{2}}(c^2\beta^2+1)^{\frac{5}{2}}}\right]\\ &\times\tilde{R}_s W^o W^{o^T}\tilde{R}_s^T = a\tilde{R}_{ss} + \frac{b}{\sigma^2}\tilde{R}_s W^o W^{o^T}\tilde{R}_s^T.\end{aligned} \quad (18)$$

Assuming $\tilde{R}_{ss}$ is positive definite, (18) can be written as

$$\nabla^2\xi(\tilde{W}) = \tilde{R}_{ss}^{\frac{1}{2}}\left\{aI + \frac{b}{\sigma^2}\tilde{R}_{ss}^{-\frac{1}{2}}\tilde{R}_s W^o W^{o^T}\tilde{R}_s^T\tilde{R}_{ss}^{-\frac{1}{2}}\right\}\tilde{R}_{ss}^{\frac{1}{2}} \quad (19)$$

where $\tilde{R}_{ss}^{1/2}$ is symmetric and nonsingular. Thus, (19) is of the form $C^T M C$ where C is nonsingular. The following result is now used [17, p. 254]: *If A is positive definite and C is nonsingular, then $C^T AC$ is also positive definite.* Thus, if M is positive definite, so is the Hessian.

The eigenvectors of

$$M = aI + \frac{b}{\sigma^2}\tilde{R}_{ss}^{-\frac{1}{2}}\tilde{R}_s W^o W^{o^T}\tilde{R}_s^T\tilde{R}_{ss}^{-\frac{1}{2}} \quad (20)$$

are $\tilde{R}_{ss}^{-1/2}\tilde{R}_s W^o$ and $N-1$ vectors orthogonal to it. Thus, M has $N-1$ eigenvalues equal to a and one eigenvalue given by $\gamma = a + (b/\sigma^2)W^{o^T}\tilde{R}_s^T\tilde{R}_{ss}^{-1}\tilde{R}_s W^o = a + b\beta^2$. M will be positive definite if all these eigenvalues are positive.

Taking the expressions of a and b from (18) and using (12) yields, after simple algebraic manipulations

$$\gamma = a + b\beta^2 = \frac{4c^4\beta^4 + 4c^2\beta^2 + 2}{(c^2\beta^2+1)^2(2c^2\beta^2+1)^{\frac{3}{2}}} > 0 \quad (21)$$

which completes the proof that the Hessian is positive definite for any finite β^2.

References

[1] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[2] P. Darlington, "Performance surfaces of minimum effort estimators and controllers," *IEEE Trans. Signal Processing*, vol. 43, pp. 536–539, Feb. 1995.
[3] P. Darlington and G. Xu, "Equivalent transfer functions of minimum output variance mean-square estimators," *IEEE Trans. Signal Processing*, vol. 39, pp. 1674–1677, July 1991.
[4] A. Stenger and W. Kellerman, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, pp. 1747–1760, 2000.
[5] C. Hansen, "Active noise control—from laboratory to industrial implementation," in *Proc. National Conf. Noise Control Engineering (NOISE-CON)*, 1997, pp. 3–38.
[6] G. Tao and P. V. Kokotovic, *Adaptive Control of Systems With Actuators and Sensor Nonlinearities*. New York: Wiley, 1996.
[7] R. J. Bernhard, P. Davies, and S. W. Kurth, "Effects of nonlinearities on system identification in active noise control systems," in *Proc. National Conf. Noise Control Engineering (NOISE-CON)*, 1997, pp. 231–236.
[8] M. H. Costa, J. C. M. Bermudez, and N. J. Bershad, "Stochastic analysis of the filtered-X LMS algorithm in systems With nonlinear secondary paths," *IEEE Trans. Signal Processing*, vol. 50, pp. 1327–1342, June 2002.
[9] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations*. New York: Wiley, 1996.
[10] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1987.
[11] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
[12] R. F. Pawula, "A modified version of Price's theorem," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 285–288, Apr. 1967.
[13] J. J. Shynk and N. J. Bershad, "Steady-state analysis of a single-layer perceptron based on a system identification model with bias terms," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 1030–1042, Sept. 1991.
[14] N. J. Bershad, P. Celka, and J. M. Vesin, "Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for Gaussian data and noisy input and output measurements," *IEEE Trans. Signal Processing*, vol. 47, pp. 675–689, Mar. 1999.
[15] O. J. Tobias, J. C. M. Bermudez, and N. J. Bershad, "Mean weight behavior of the filtered-X LMS algorithm," *IEEE Trans. Signal Processing*, vol. 48, pp. 1061–1075, Apr. 2000.
[16] M. H. Costa, J. C. M. Bermudez, and N. J. Bershad, "Stochastic analysis of the LMS algorithm with a saturation nonlinearity following the adaptive filter output," *IEEE Trans. Signal Processing*, vol. 49, pp. 1370–1387, July 2001.
[17] G. Strang, *Linear Algebra and its Applications*. New York: Academic Press, 1980.