By (10), the right side of (11) contains the term

$$\sum_{k=1}^{n-1} \lambda_{k+1}^T \Psi_k \frac{\partial \Sigma_k}{\partial \theta_j} H_k^T B_k^{-1} \nu_k = \text{tr} \left(\sum_{k=1}^{n-1} \Psi_k \frac{\partial \Sigma_k}{\partial \theta_j} H_k^T B_k^{-1} \nu_k \lambda_{k+1}^T \right).$$

Since Σ_k is symmetric

$$\text{tr} \left(\sum_{k=1}^{n-1} \Psi_k \frac{\partial \Sigma_k}{\partial \theta_j} H_k^T B_k^{-1} \nu_k \lambda_{k+1}^T \right) = \text{tr} \left(\sum_{k=1}^{n-1} \frac{\partial \Sigma_k}{\partial \theta_j} U_k \right) \tag{13}$$

where

$$U_k = \frac{1}{2} (H_k^T B_k^{-1} \nu_k \lambda_{k+1}^T \Psi_k + \Psi_k^T \lambda_{k+1} \nu_k^T B_k^{-1} H_k).$$

The right side of (13) can also be rewritten by applying a matrix adjoint result [6]. If $S_{j,k+1} = A_k S_{jk} A_k^T + D_{jk}$, $k = 1, \cdots, n$, $1 \le j \le q$, S_{j1} given and symmetric, and if

$$L = \text{tr} \left(\sum_{k=1}^{n} S_{jk} C_k \right), \quad C_k \text{ symmetric, then}$$

$$L = \text{tr} \left(\Lambda_1 S_{j1} + \sum_{k=1}^{n-1} \Lambda_{k+1} D_{jk} \right)$$

where

$$\Lambda_k = A_k^T \Lambda_{k+1} A_k + C_k, \quad k = n-1, \cdots, 1,$$

$$\Lambda_n = C_n.$$

When $S_{jk} = \partial \Sigma_k / \partial \theta_j$, it follows from (6) that

$$\frac{\partial \Sigma_{k+1}}{\partial \theta_j} = \Psi_k \frac{\partial \Sigma_k}{\partial \theta_j} \Psi_k^T + \Omega_{kj} + \Omega_{kj}^T$$

where

$$\Omega_{kj} = \frac{\partial F_k}{\partial \theta_j} \Sigma_k \Psi_k^T - F_k K_k \frac{\partial H_k}{\partial \theta_j} \Sigma_k \Psi_k^T + \frac{1}{2} \frac{\partial Q_k}{\partial \theta_j} + \frac{1}{2} F_k K_k \frac{\partial R_k}{\partial \theta_j} K_k^T F_k^T.$$

Hence, applying the matrix adjoint result we obtain

$$\text{tr} \left(\sum_{k=1}^{n-1} \frac{\partial \Sigma_k}{\partial \theta_j} U_k \right) = \text{tr} \left(\Lambda_1 \frac{\partial \Sigma_1}{\partial \theta_j} \right) + 2 \text{ tr} \left(\sum_{k=1}^{n-1} \Lambda_{k+1} \Omega_{kj} \right) \tag{14}$$

where

$$\Lambda_k = \Psi_k^T \Lambda_{k+1} \Psi_k + U_k \quad (k = n-1, \cdots, 1)$$

$$\Lambda_n = 0. \tag{15}$$

From (8), (11), and (14), the derivative $\partial J / \partial \theta_j$ can be written as

$$\frac{\partial J}{\partial \theta_j} = -\lambda_1^T \frac{\partial \hat{x}_1}{\partial \theta_j} - \sum_{k=1}^{n-1} \lambda_{k+1}^T \omega_{kj} - \text{tr} \left(\Lambda_1 \frac{\partial \Sigma_1}{\partial \theta_j} \right) - 2 \text{ tr} \left(\sum_{k=1}^{n-1} \Lambda_{k+1} \Omega_{kj} \right) - \text{tr} \left(\sum_{k=1}^{n} \hat{x}_k \nu_k^T W \frac{\partial H_k}{\partial \theta_j} \right)$$

where, for convenience,

$$\omega_{kj} = \frac{\partial F_k}{\partial \theta_j} (\hat{x}_k + K_k \nu_k) + \Psi_k \Sigma_k \frac{\partial H_k^T}{\partial \theta_j} B_k^{-1} \nu_k - F_k K_k \left[\frac{\partial H_k}{\partial \theta_j} (\hat{x}_k + K_k \nu_k) + \frac{\partial R_k}{\partial \theta_j} B_k^{-1} \nu_k \right]$$

the λ_k and Λ_k's being given by (12) and (15), respectively.

IV. Summary

A new expression has been obtained for the derivatives of the prediction error cost function for time-varying linear dynamic systems. The main advantage of this new expression is that it requires fewer computations to obtain the gradient than does the straightforward approach via sensitivity equations.

Acknowledgment

The author wishes to thank Dr. J. L. Maryak of the Applied Physics Laboratory for several useful suggestions.

References

[1] P. E. Caines, "Prediction error identification methods for stationary stochastic processes," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 500–505, Aug. 1976.

[2] B. D. O. Anderson, J. B. Moore, and R. M. Hawkes, "Model approximations via prediction error identification," *Automatica*, vol. 14, pp. 615–622, Nov. 1978.

[3] L. Ljung and A. J. M. Van Overbeek, "Validation of approximate models obtained from prediction error identification," in *Proc. Int. Fed. Automat. Contr. 7th Trien. World Cong.*, Helsinki, Finland. London, England: Pergamon, 1978, pp. 1899–1906.

[4] K. J. Åström, "Maximum likelihood and prediction error methods," *Automatica*, vol. 16, pp. 551–574, 1980.

[5] T. Söderström and P. Stoica, "On criterion selection and noise model parametrization for prediction error identification methods," *Int. J. Contr.*, vol. 34, pp. 801–811, Oct. 1981.

[6] N. R. Sandell, Jr., and K. I. Yared, "Maximum likelihood identification of state space models for linear dynamic systems," Electron. Syst. Lab., M.I.T., Cambridge, MA, Rep. ESL-R-814, 1978.

[7] G. A. Van Zee, "Gradient computation in prediction error identification of linear discrete-time systems," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 738–739, June 1982.

[8] D. A. Wilson and A. Kumar, "Derivative computations for the log-likelihood function," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 230–232, Feb. 1982.

On the Identification of Polynomial Input–Output Differential Systems

A. E. PEARSON AND F. C. LEE

***Abstract*—A least-squares parameter identification technique is formulated for a class of deterministic nonlinear systems modeled by polynomial input–output differential equations. The basis of the technique is Shinbrot's method of moment functionals using trigonometric modulating functions. Given the input–output data over a single finite time interval for a one-shot estimate, or over a sequence of finite time intervals for sequential least squares, the underlying computations utilize a fast fourier transform algorithm on polynomials of the data without the need for estimating unknown initial or boundary conditions at the start of each finite time interval.**

I. Introduction

The parameter identification of deterministic nonlinear systems modeled by polynomial type differential equations can be undertaken by the Bellman–Kalaba quasi-linearization technique [1], [2], or by finite-dimensional hill climbing techniques after approximating all signals with

Manuscript received March 15, 1984; revised November 8, 1984. Paper recommended by Past Associate Editor, B. Friedlander. This work was supported in part by the National Science Foundation under Grant ECS-8111219 and in part by the Air Force Office of Scientific Research under Grant AFOSR-82-0230.

A. E. Pearson is with the Division of Engineering, Brown University, Providence, RI 02912.

F. C. Lee is with the Institute of Control Engineering, National Chiao Tung University, Hsinchu, Taiwan, Republic of China.

piecewise constant Walsh-type orthogonal functions [3]. The former entails the iterative solution to a nonlinear two point boundary value problem while the latter approach gives up the differential equation model in the very beginning. In both approaches unknown initial conditions have to be estimated along with the system parameters for time limited data. A different approach is taken here for a restricted class[1] of polynomial differential systems which retains the continuous-time format while avoiding the necessity to deal with all unknown boundary (initial) conditions for input-output data collected over a fixed finite time interval, or over a sequence of finite time intervals $[t_i, t_{i+1}]$, each of duration T. This is accomplished by projecting the data down into a suitable subspace via Shinbrot's "method of moment functionals" [5] using finite sums of commensurable sinusoids as the modulating functions. As developed by the authors in [6] for linear systems, the underlying computations entail calculating a fixed number of Fourier series coefficients for the data on each $[0, T]$ interval.[2] It is known that these coefficients can be determined with a high degree of accuracy using the discrete Fourier transform (DFT) which, in turn, can be evaluated very efficiently using the fast Fourier transform (FFT) algorithm. These computational aspects, together with the results of a simulated example, will be discussed following the formulation of the problem.

II. Formulation

Let $[u(t), y(t)]$ denote an input-output pair for a single input, single output system which is observed free of significant measurement noise over a fixed finite time interval $[0, T]$: $0 \le t \le T$, or over a sequence of time intervals $[t_i, t_{i+1}]$, $i = 0, 1 \cdots$, each of duration T. It is assumed that the data are bounded and piecewise continuous on every finite time interval. Let p denote the differential operator d/dt so that $p^2 = d^2/dt^2$, etc. Corresponding to *a priori* integers (n, m) the class of models relating $u(t)$ and $y(t)$ is defined by the polynomial input-output differential operator equation

$$p^n y(t) + \sum_{i=1}^{n}\sum_{j=0}^{m}\sum_{k=0}^{m} a_i(j, k) p^{n-i}[u(t)]^j[y(t)]^k = 0$$

$$0 \le t \le T, \; a_i(0, 0) = 0, \qquad i = 1 \cdots n. \tag{1}$$

The $a_i(j, k)$ represent parameters—a total of $mn(m + 2)$ in the general case—which are to be determined by a least-squares technique without the need for estimating unknown initial conditions at the start of each $[t_i, t_{i+1}]$ interval.[3]

Defining the column n-vector parameters $\{\alpha, \beta, \eta(j), \nu(k), \gamma(j, k)\}$ by their components

$$\alpha_i = a_i(0, 1) \quad \eta_i(j) = a_i(j, 0), \; 2 \le j \le m$$
$$\beta_i = a_i(1, 0) \quad \nu_i(k) = a_i(0, k), \; 2 \le k \le m$$
$$\gamma_i(j, k) = a_i(j, k), \; 1 \le j, k \le m$$
$$i = 1, 2, \cdots, n, \tag{2}$$

it can be readily verified that a set of state equations equivalent to the model (1) is specified by

$$\dot{x} = \left[\begin{array}{c|c} -\alpha & \begin{matrix} & \\ & I_{n-1} \\ & \end{matrix} \\ \hline \vdots & \\ 0 & \end{array}\right] x - \beta u - \sum_{j=2}^{m} \eta(j)[u]^j - \sum_{k=2}^{m} \nu(k)[x_1]^k$$
$$- \sum_{j=1}^{m}\sum_{k=1}^{m} \gamma(j, k)[u]^j[x_1]^k \tag{3}$$

where $y = x_1$ is the first component of the column n-vector x. Within the context of analytic differential systems of the form

$$\dot{x} = f(x, u)$$
$$y = g(x)$$

where f and g are polynomial functions of their arguments, the model (3) is seen to represent the special class wherein the output function g is linear in x, i.e., $g(x) = c'x$ for some row vector c', and the nonlinear terms in f are polynomials in the scalar pair $(c'x, u)$. It is easy to specialize certain well-known models to the above form such as the Duffing and Van der Pol equations.

Shinbrot's method of moment functionals is a classical technique for converting a differential equation to an algebraic equation in the parameters by the use of so-called modulating functions. As introduced by Shinbrot [5], $\phi(t)$ is a modulating function of order n relative to a fixed time interval $[0, T]$ if it is sufficiently smooth and satisfies the end point conditions

$$\phi^{(i)}(0) = \phi^{(i)}(T) = 0$$
$$i = 0, 1, \cdots, (n-1) \tag{4}$$

where $\phi^{(i)}(t)$ means $p^i\phi(t)$. The significance of this property for the model (1) relates to the fact that if (1) is multiplied by $\phi(t)$ and integrated over $[0, T]$, the result is the functional equation

$$(-1)^n \int_0^T y(t)\phi^{(n)}(t)\, dt$$
$$+ \sum_{i=1}^{n}\sum_{j=0}^{m}\sum_{k=0}^{m} a_i(j, k)(-1)^{n-i} \int_0^T [u(t)]^j[y(t)]^k \phi^{(n-i)}(t)\, dt = 0. \tag{5}$$

This follows using integration by parts n times and noting the end point conditions (4). Moreover, if $\{\phi_i(t)\}$, $i = 1, 2, \cdots K$, is a set of linearly independent modulating functions, a vector algebraic equation results which can be used to obtain a least-squares estimate of the parameters.

As pointed out in [6] the above idea has been pursued by several investigators using a variety of modulating functions such as Hermite polynomials and splines. However, the computational burden associated with these functionals on the data will generally be significant unless a "fast algorithm" is available. Such is the case for modulating functions comprised of linear combinations of commensurable sinusoids since integrals like

$$\int_0^T [u(t)]^j[y(t)]^k \; [\sin l\omega_0 t \text{ or } \cos l\omega_0 t]\, dt$$

$$\omega_0 = \frac{2\pi}{T}, \qquad l = 0, 1, 2, \cdots, L$$

can be efficiently evaluated by an FFT algorithm especially for large L. With this in mind, let $f(t)$ denote the $(2L + 1)$ column vector of sinusoids defined by

$$f(t) = \text{col } [1; \cos \omega_0 t, \sin \omega_0 t; \cos 2\omega_0 t, \sin 2\omega_0 t; \cdots, \cos L\omega_0 t,$$
$$\sin L\omega_0 t]$$
$$0 \le t \le T, \qquad \omega_0 = \frac{2\pi}{T} \tag{6}$$

where the integer L must be chosen at least as large as $n/2$ in order to satisfy (4). Roughly speaking, the pair (L, T) will be selected by the user so that $2L$ is somewhat larger than the system model order n and T is sufficiently long to assure that $\omega_0 = 2\pi/T$ is small enough to resolve the spectral components of the data. This will be discussed further in the next section.

Using the (off line) procedure outlined in [6], let C denote the $(2L + 1$

[1] The various models commonly used in nonlinear systems identification can be found in survey articles such as Haber and Keviczky [4].

[2] An alternative projection method which could be brought to bear on the same class of models is the "projected integral equation error" technique presented in [7]. However, it is believed that the use of the FFT algorithm makes the approach of this paper computationally superior to the formulation in [7].

[3] It will be tacitly assumed that any candidate values assigned to the parameters $a_i(j, k)$ are such that (1) possesses bounded solutions over the time intervals of interest, i.e., no finite escape times.

$-n) \times (2L + 1)$ matrix constructed such that $\Phi(t)$ defined by[4]

$$\Phi(t) = Cf(t), \qquad 0 \le t \le T \tag{7}$$

satisfies the end point conditions

$$\Phi^{(i)}(0) = \Phi^{(i)}(T) = 0, \qquad i = 0, 1, \cdots (n-1). \tag{8}$$

Thus, $\Phi(t)$ is a $(2L + 1 - n)$ dimensional vector valued modulating function of order n in which the derivatives $p^i\Phi(t) = \Phi^{(i)}(t)$ have the representation [cf. (6)]

$$(-1)^i p^i \Phi(t) = CD^i f(t), \qquad i = 0, 1 \cdots \tag{9}$$

where D is the block diagonal matrix defined by

$$D = \omega_0 \begin{bmatrix} 0 & & & & \\ & \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} & & & \\ & & \begin{matrix} 0 & 2 \\ -2 & 0 \end{matrix} & & \\ & & & \ddots & \\ & & & & \begin{matrix} 0 & L \\ -L & 0 \end{matrix} \end{bmatrix} \tag{10}$$

and D^0 is defined as the identity matrix. Then multiplying the model equation (1) with $\Phi(t)$, integrating over $[0, T]$, and using integration-by-parts n times while noting (8), there results the vector analog of (5):

$$CD^n Y + \sum_{i=1}^{n} \sum_{j=0}^{m} \sum_{k=0}^{m} a_i(j, k) CD^{n-i} Z(j, k) = 0 \tag{11}$$

where $Z(j, k)$ and Y are (finite) Fourier coefficient vectors defined by

$$Z(j, k) = \int_0^T [u(t)]^j [y(t)]^k f(t)\, dt, \qquad 0 \le j, k \le m$$

$$Y = Z(0, 1). \tag{12}$$

Equation (11) can be written in the standard form for least squares by defining the column block partitioned parameter vector θ [cf. (2)]

$$\theta = \text{col } [\alpha, \beta, \eta(2) \cdots \eta(m), \upsilon(2) \cdots \upsilon(m), \gamma(1, 1) \cdots \gamma(m, m)] \tag{13}$$

and the coefficient M_{jk} (each of dimension $(2L + 1) \times n$) according to the partitions

$$M_{jk} = [D^{n-1} Z(j, k), D^{n-2} Z(j, k) \cdots Z(j, k)] \tag{14}$$

such that the product $M\theta$ results in the vector sum [cf. (2) and (3)]

$$M\theta = M_{01}\alpha + M_{10}\beta + \sum_{j=2}^{m} M_{j0}\eta(j) + \sum_{k=2}^{m} M_{0k}\upsilon(k) + \sum_{j=1}^{m} \sum_{k=1}^{m} M_{jk}\gamma(j, k). \tag{15}$$

That is to say, matrix M is partitioned into row blocks

$$M = \text{Row } [M_{01}, M_{10}, M_{20} \cdots M_{m0}, M_{02} \cdots M_{0m}, M_{11} \cdots M_{mm}] \tag{16}$$

comformably with the partitioning in θ such that (15) holds. With these definitions, (11) is equivalent to

$$CD^n Y + CM\theta = 0. \tag{17}$$

[4] It can be shown that C has full rank and that determining C involves solving $(2L + 1 - n)$ Vandermonde type matrix equations. Further details can be found in [11].

The normal equation for (17) is given by

$$M'C'CM\theta = M'C'CD^n Y \tag{18}$$

and a unique solution for a one-shot least-squares estimate of θ is obtained if and only if the Gram matrix $M'C'CM$ has full rank. Although the matrix C has full rank $(2L + 1 - n)$, it is not enough to assert that a unique solution to (18) exists if M has full rank. Hence, the uniqueness of the one-shot least-squares estimate is predicated on the condition that CM has full rank equal to $mn(m + 2)$. In turn, this implies the inequality $(2L + l - n) \ge mn(m + 2)$ since C has dimension $(2L + 1 - n) \times (2L + 1)$. Therefore, the one-shot least-squares estimate has the potential for being well posed only if $2L \ge [mn(m + 2) + n - 1]$. This supercedes the basic inequality $2L \ge n$ which arises from the end point conditions (8).

The above equations pertain to data observed over a single $[0, T]$ time interval. In the case of data observed over sequential time intervals $[t_i, t_{i+1}]$, $i = 0, 1 \cdots$, each of duration T, (17) is replaced by

$$CD^n Y(i) + CM(i)\theta = 0 \tag{19}$$

where $Y(i)$ and $M(i)$ are computed from (12) and (14) for the data collected over each T-interval. Standard sequential least-squares theory for deterministic discrete models can then be applied to (19) in constructing a recursive solution $\theta(i)$.[5]

III. Some Computational Considerations

A. Choice of (L, T)

Borrowing from frequency domain ideas for linear systems, the choice of the time interval $[0, T]$ can be based on the heuristic notion that since $\omega_0 = 2\pi/T$ is essentially the resolving frequency, T should be large enough to distinguish the characteristic modes of the system. Likewise, the selection of the ratio L/T can be guided by the consideration that $L\omega_0 = L2\pi/T$ be comparable to the system bandwidth. Although "characteristic mode" and "system bandwidth" are not well-defined concepts for nonlinear systems, their intuitive meanings bear some relevance to this discussion. Also, since the modulating functions act as a filter on the data through the finite Fourier series coefficients, L should not be too large as to give undue emphasis to whatever high-frequency measurement noise might be present in the data.

B. Computing Z(j, k) Via FFT Techniques

The major computational burden in setting up the least-squares identification of (1) will be the determination of the $Z(j, k)$ in (12) at each stage. Let $z(t)$ denote a typical function from the $m(m + 1)$ set of functions $\{[u(t)]^j [y(t)]^k, 0 \le j, k \le m\}$, $0 \le t \le T$. Then (6) and (12) imply determining the following integrals (complex form) for each $z(t)$:

$$\int_0^T z(t) e^{jl\omega_0 t}\, dt, \qquad l = 0, 1 \cdots L. \tag{20}$$

Although the real and imaginary parts of the above integral can be evaluated by passing $z(t)$ through a bank of appropriately tuned harmonic oscillators [6], greater flexiblity is offered by using well-known digital approximations. Thus, for example, if uniform sampling of $z(t)$ is used to generate N samples $z_i = z(ih)$, $h = T/N$, $i = 0, 1, \cdots (N - 1)$, the standard parabolic rule yields

$$\int_0^T z(t) e^{jl\omega_0 t}\, dt = \frac{h}{3} \left[z_0 + z_N + 4 \sum_{i=1,3\cdots}^{N-1} z_i W^{li} + 2 \sum_{i=2,4\cdots}^{N-2} z_i W^{li} \right] + o(h^4) \tag{21}$$

[5] See, for example, Mendel [8] for convergence theorems and a discussion of various practical problems, such as initialization and roundoff errors, which attend all sequential least-squares formulations.

where $W = e^{j2\pi/N}$ and $o(\cdot)$ is the order of the error as a function of the sampling interval h. Assuming N is a power of 2, the usual FFT algorithm can be used to evaluate the DFT of the quantity in brackets on the right-hand side of the above approximation yielding the Fourier coefficients for $l = 0, 1 \cdots (N-1)$, i.e.,

$$Z = \frac{h}{3} \text{ FFT } [(z_0 + z_N),\ 4z_1,\ 2z_2,\ 4z_3,\ 2z_4 \cdots 4z_{N-1}]. \tag{22}$$

The computational savings of this algorithm for large N are well known; specifically, $\log_2 N/N$. However, a special FFT-type algorithm can be devised in consideration of the fact that only L Fourier coefficients are needed in the computation and the efficiency of such an algorithm is $\log_2 L/L$. As pointed out by a reviewer, this represents a kind of FFT "pruning" discussed in Markel [9].

C. A Simulated Example

The following bilinear control system model for nuclear fission is discussed in Mohler [10, Section 4.1]

$$l\dot{y} = (u - \beta)y + l\lambda\bar{c}$$

$$l\dot{\bar{c}} = \beta y - l\lambda\bar{c}.$$

Eliminating the "average precursor population" variable $\bar{c}(t)$, the preceding equations in the "neutron population" $y(t)$ become

$$p^2y(t) + \left(\lambda + \frac{\beta}{l}\right)py(t) = \frac{1}{l}(p + \lambda)[u(t)y(t)]$$

where $u(t)$ is the control reactivity and (λ, β, l) are parameters characterizing the nuclear fission process. Using the sinusoidal forcing function

$$u(t) = 2 + \sin\left(t + \frac{\pi}{4}\right) + \sin\left(2t + \frac{\pi}{4}\right)$$

several simulation runs were made over a $[0, 2\pi]$ time interval, i.e., $\omega_0 = 2\pi/T = 1$, for a one-shot least-squares estimate of the following parameters:

$$\theta_1 = \lambda + \frac{\beta}{l} = 2.0$$

$$\theta_2 = -\frac{\lambda}{l} = -1.0$$

$$\theta_3 = -\frac{1}{l} = -1.0.$$

Fig. 1 shows the output data $y(t)$ for noise-free conditions (a) and a run (b) in which white Gaussian noise was added to the data resulting in a noise-to-signal ratio of approximately 5 percent. Notice that the system is apparently unstable for the above chosen parameter values and control reactivity.

Table I lists the estimated parameter values for the different runs using the modulating function frequencies $\{0, 1, 2, 3, 4\}$, i.e., $L = 4$.[6] This means that $2L + 1 - n = 7$ algebraic equations in (17) were used to form the normal equation (18) for each run. No difficulties were encountered in solving the normal equation (18) for any of the runs. The two columns under each parameter estimate in the table give the results for two different discrete approximations in computing the Fourier coefficients (21) for each $z(t) \in \{u(t), y(t), u(t)y(t)\}$, i.e., the FFT orders were $N = 128$ and $N = 256$ in (22).[7] The results show that good accuracy in the parameter estimates is attained under ideal noise free

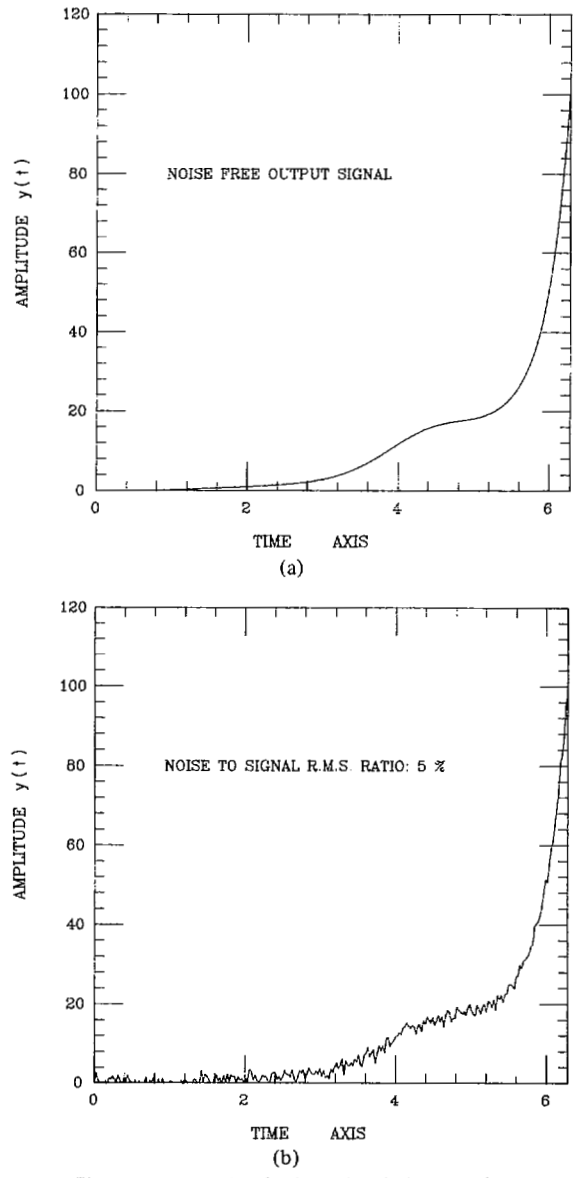

Fig. 1. Response data for the nuclear fission example.

TABLE I
ESTIMATED PARAMETER VALUES FOR THE NUCLEAR FISSION EXAMPLE

	$\hat{\theta}_1$		$-\hat{\theta}_2$		$-\hat{\theta}_3$		$\|\|\Delta\theta\|\|$	
Order FFT / Run	128	256	128	256	128	256	128	256
0	1.939		0.994		0.965		2.7%	
1*	1.223	1.898	0.982	0.986	0.493	0.938	36.9%	4.7%
2*	2.110	1.924	0.447	0.947	1.088	1.010	6.8%	3.8%
3*	1.220	2.129	1.019	0.942	0.484	1.138	37.4%	9.4%
4*	1.625	1.921	0.936	0.993	0.809	0.954	15.9%	3.5%

*Runs 1-4 include additive noise in the data with an RMS N/S ratio $\simeq$ 5%.

conditions (Run 0) using the relatively coarse DFT order ($N = 128$), but that a similar degree of accuracy could be maintained under the more realistic measurement noise case (Runs 1–4) only by increasing the order of the DFT ($N = 256$). Limiting the highest modulating function frequency by choosing the value $L = 4$ served to filter out the higher frequencies in the white measurement noise, i.e., increasing L will decrease the estimation accuracy in the noisy case. At the same time, reducing L will result in fewer algebraic equations for the least-squares estimate. A rule of thumb is to choose L such that $(2L + 1 - n)$ is approximately double the unknown parameters for a one-shot estimate

[6] The modulating functions themselves were obtained by separately subjecting linear combinations of the functions in the two sets $\{\sin it, 1 \le i \le 4\}$ and $\{\cos it, 0 \le i \le 4\}$ for $t \in [0, 2\pi]$ to the end point constraints (4). As mentioned earlier, this is an off line calculation involving the solution to Vandermonde type matrix equations yielding the matrix C in (7).

[7] The IMSL Library was used to provide the integration routine (DVERK) for generating the "continuous" data and as the source for an FFT algorithm to compute the DFT's of the "sampled" data.

and to choose T such that $L2\pi/T$ is comparable to the system bandwidth (if known).

IV. Conclusions

Choosing commensurable sinusoids as modulating functions in the Shinbrot method of moment functionals, it has been shown how the least-squares identification of polynomial input–output differential systems can be formulated in a way that utilizes the computationally efficient FFT algorithm at each stage while avoiding the necessity to estimate unknown initial conditions for time limited data. In addition to the order of the system model and the number of parameters to be identified, the choice in modulating functions can be based to some extent on noise rejection and the heuristic notion of "system bandwidth." An interesting problem for future investigation is determining an optimal set of Fourier-based modulating functions which minimizes some measure of the error in the parameter estimates for noisy measurements with specified spectral characteristics. Another problem is determining conditions on the input data that guarantee uniqueness of the least-squares estimate.

References

[1] R. E. Bellman and R. E. Kalaba, *Quasilinearization and Nonlinear Boundary Value Problems.* New York: American Elsevier, 1965.

[2] R. Kalaba and K. Spingarn, *Control, Identification and Input Optimization.* New York: Plenum, 1982.

[3] G. P. Rao, *Piecewise Constant Orthogonal Functions and Their Application to Systems and Control.* New York: Springer-Verlag, 1983.

[4] R. Haber and L. Keviczky, "Identification of nonlinear dynamic systems," in *Identification and System Parameter Estimation, Part 1,* N. S. Rajbman, Ed. Amsterdam, The Netherlands: North-Holland, 1978, pp. 79–126.

[5] M. Shinbrot, "On the analysis of linear and nonlinear systems," *Trans. ASME,* vol. 79, pp. 547–552, Apr. 1957.

[6] A. E. Pearson and F. C. Lee, "Time limited identification of continuous systems using trigonometric modulating functions," in *Proc. of 3rd Yale Workshop on Applications of Adaptive Syst.*, New Haven, CT, 1983, pp. 168–173.

[7] A. E. Pearson, "Nonlinear system identification with limited time data," *Automatica,* vol. 15, pp. 73–84, Jan. 1979.

[8] J. M. Mendel, *Discrete Techniques of Parameter Estimation.* New York: Marcel-Dekker, 1973.

[9] J. D. Markel, "FFT pruning," in *IEEE Trans. Audio. Electroacoust.*, vol. AU-19, pp. 305–311, Dec. 1971.

[10] R. R. Mohler, *Bilinear Control Processes.* New York: Academic, 1973.

[11] F. C. Lee, "Time limited identification of continuous systems using the modulating function method," Ph.D. dissertation, Division Eng., Brown Univ., Providence, RI, 1984.

Nonparametric Kernel Algorithm for Recovery of Functions from Noisy Measurements with Applications

ALEXANDER A. GEORGIEV

Abstract—**This note presents a kernel algorithm for recovery of a regression function from noisy data. Conditions are provided that assure pointwise convergence in the mean square and almost sure senses. An application to a class of linear system identification problems is discussed.**

I. Introduction

There are many situations in system identification when a nonlinear memoryless system $g(x)$ is estimated from observations $(x_1, Y_1), (x_2, Y_2), \cdots, (x_n, Y_n)$, where the design points x_i are from a real interval [0, 1] over which observations Y_i are taken, satisfying

$$Y_i = g(x_i) + Z_i. \tag{1}$$

Assume that the points x_i are selected in some arbitrary fashion by the experimenter and are not necessarily equally spaced. The errors $Z_1, \cdots, Z_n$ are mean-zero independent identically distributed random variables with finite variance σ^2. It is interesting for many purposes to estimate the system g, for example, to predict g at a particular point x and to describe and illustrate the relationship between the design points x_i and the observations Y_i.

In the finite-dimensional parameter case, when the function g is known except for a set of parameters, the least square or the maximum likelihood methods are commonly used [7], [16].

In this note we propose a nonparametric kernel algorithm for recovery of the function g from noisy measurements. The "nonparametric" property refers to the absence of a finite-dimensional continuous parametrization of the space of functions containing g.

Assume that the input signals x_i satisfy, without loss of generality, the order condition

$$x_1 < x_2 < \cdots < x_n. \tag{2}$$

Let $x_o = 0$ and $x_{n+1} = 1$. The considered kernel algorithm for recovery of $g(x)$ is

$$g_n(x) = \sum_{i=1}^{n} Y_i K\left(\frac{x - x_i}{h_n}\right)\frac{(x_i - x_{i-1})}{h_n} \tag{3}$$

where K (so-called kernel) is a bounded nonnegative function on the real line and $\{h_n\}$ is a sequence of positive real numbers. The estimate (3) can be regarded as the appropriate average of observations in a neighborhood of the point under consideration. For example, one might consider

$$\hat{g}_n(x) = \frac{1}{nh_n} \sum_{s(x)} Y_i \tag{4}$$

where $s(x) = \{i : |x - x_i| \leq h_n\}$. The procedure $\hat{g}_n$ is obtainable from (3) for the window kernel

$$K(u) = \begin{cases} 1, & |u| \leq 1 \\ 0, & |u| > 1 \end{cases} \tag{5}$$

and for the x_i's equally spaced, i.e., $x_i - x_{i-1} = 1/n$.

The estimate g_n was introduced by Priestley and Chao [11] and has been studied by Benedetti [1], Gasser and Müller [3], Schuster and Yakowitz [17], as well as Cheng and Lin [2]. Another method, the "nearest neighbor method," for the recovery of function g has been introduced by Greblicki [9] and an orthogonal series method has been examined by Rutkowski [14].

The main object of this paper is to introduce control engineers to some techniques afforded by nonparametric methodology. In Section II, the convergence theorems are stated and new results are discussed. Applications to linear dynamical system identification are given in Section III. Finally, some remarks about nonparametric identification procedures are made.

II. The Convergence of the Algorithm

In this section, we establish the sufficient conditions for consistency of the algorithm (3). The proofs of the following results are given in the Appendix.

Lemma (Bias):

Assume that $g(x)$ is a bounded function and K is a continuous probability density function such that $K(u)$ is nonincreasing for $u > 0$, and nondecreasing for $u < 0$. Let

$$\delta_n = \max_i (x_i - x_{i-1}) = 0(n^{-1}). \tag{6}$$

Manuscript received December 19, 1983; revised July 30, 1984. This paper is based on a prior submission of February 28, 1983.

The author is with the Institute of Engineering Cybernetics, Technical University of Wroclaw, Wroclaw, Poland.

0018-9286/85/0800-0782$01.00