Two Papers on Bilinear Systems:

a) A Study of Second- and Third-order Spectral Procedures and Maximum Likelihood Identification of a Bilinear System


David R. Brillinger

Statistics Department
University of California

Technical Report No. 199
March 1989

*Research supported by National Science Foundation Grant DMS-8316634.

Department of Statistics
University of California
Berkeley, California
A Study of Second- and Third-Order Spectral Procedures and Maximum Likelihood in the Identification of a Bilinear System

DAVID R. BRILLINGER

Abstract- This paper describes a comparative investigation of two spectral moment-based identification procedures with each other and with the method of maximum likelihood. The case considered is that of a bilinear system. Analyses making use of second- and third-order moments are presented. Inputs investigated are pulse, pulse-pair and stochastic. Some of the relative advantages and disadvantages of each estimation procedure are mentioned. Advantages of the maximum likelihood approach include the availability of expressions for standard errors and efficiency in an asymptotic sense. It seems that each procedure has a role to play. One concern of the analysis is the detection of coupled frequencies.

I. INTRODUCTION

Historically the way that many statistical techniques have evolved is via an initial method of moments approach, later followed by full modelling and maximum likelihood (and more recently with the latter followed by robust/resistant variants.) These approaches are associated with the names of K. Pearson, R. A. Fisher (and J. W. Tukey) respectively. Time series analysis seems to be following along this same road. In this paper the intent is to consider the case of a particular nonlinear time series system and to contrast method of moments and maximum likelihood techniques in that case. (Because the data motivating the work is of high quality, the robust/resistant problem will not be addressed.) The system studied is bilinear and solvable. Because explicit expressions are available, the bilinear provides a good test case for the various higher-order spectral procedures being developed.

Spectral analysis provides a direct way to study systems with resonant properties and to gain insight into their structure. Fourier transforms of a single variable have long been employed; however certain information can escape such an analysis. Consequently there has been a move to employing higher-order Fourier transforms and spectra. Such transforms can resolve overlapping spectra lines, can enhance sensitivity and can provide information unavailable in first-order analyses. The resonances of the bilinear system studied are basic to its understanding.

Bilinear systems occur in important physical situations. Motivating the work of this paper is nuclear magnetic resonance (NMR) spectroscopy. The basic equations of the NMR phenomenon are bilinear. In operation a substance is placed in a strong magnetic field and then further perturbed by a radio-frequency field, \(X(t)\). A fluctuating voltage signal, \(Y(t)\), is recorded. In NMR spectroscopy in simplest form, the empirical Fourier transform of output data, \(Y(t)\), is computed and its amplitude examined for peaks. The locations of the peaks give important information concerning the composition of the substance. More recently quadratic and higher-order analyses based on Fourier transforms have been employed, see Billmich (1985). In Brillinger (1985) a maximum likelihood analysis making use of the Kalman filter was introduced. This present paper is concerned with a comparison of these various procedures. It is

The author is with the Department of Statistics, University of California, Berkeley, CA 94720.
to be expected that both the inclusion of higher-order information and the move to maximum likelihood will lead to improved efficiency.

The form of bilinear system to be considered consists of the following system and measurement equations

\[ \frac{dS(t)}{dt} = a + AS(t) + BS(t)X(t) \]  
\[ Y(t) = \text{Re}(e^{cS(t)}) \]

where \( X \) and \( Y \) represent the system's (scalar) input and output respectively, where \( S \) is a state vector and where \( a, A, B, c \) are matrices of appropriate dimensions. It will be convenient to allow the entries of \( S(t) \), \( a, A, B, c \) to be complex-valued. An important concern of this present work is the estimation of these entries and of functions of them. In the NMR case the state vector has an interpretation as a magnetization vector and \( X(t) \) as a magnetic field acting on \( S \). Expression (1) is the equation of motion. Its form is the reason for calling the system bilinear.

In the case of NMR spectroscopy the equations (1) are known as the Bloch equations. In one elementary case, see Knight and Kaiser (1982), the matrices appearing have the form

\[
\begin{align*}
a &= \begin{bmatrix} 0 & 0 \\ 0 & M_0 T_1 \end{bmatrix} \\
A &= \begin{bmatrix} -1/T_2 & \Delta \\ -\Delta & -1/T_2 \end{bmatrix} \\
B &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}
\end{align*}
\]

with \( M_0 \) denoting an equilibrium value, with \( T_1 \) and \( T_2 \) decay times and with \( \Delta \) a frequency offset. In NMR the system may be probed with pulses, pulse-pairs, pseudo-random binary noise and Gaussian white noise amongst other choices, see e.g. Blümich (1985).

A solution for the equations (1) may be set down as follows. Let \( P \) denote the solution of

\[ \frac{dP(t,u)}{dt} = (A + BX(t))P(t,u), \quad P(u,u) = I \]

(\( P \) is the so-called matricant.) The solution of (1) may now be written

\[ S(t) = P(t,t_0)S(t_0) + \int_{t_0}^{t} P(t,u)du)a \]

see Gantmacher (1959), Chapter XIV.

An alternate way to develop a solution is the following: from equation (1) it is apparent that the steady state of \( S \), when there is no input, is \(-A^{-1}a\). This suggests transforming to the new vector \( R(t) = S(t) + A^{-1}a \). The equation (1) now becomes

\[ \frac{dR(t)}{dt} = AR(t) + BR(t)X(t) + CX(t) \]

with \( C = -BA^{-1}a \). Taking \( R(-\infty) = 0 \) and integrating (6) leads to

\[ R(t) = \int_{-\infty}^{t} e^{A(t-s)}CX(s)ds + \int_{-\infty}^{t} e^{A(t-s)}BR(s)X(s)ds \]

supposing the integrals are finite. Substituting for \( R(s) \) on the right-hand side of (7) leads to

\[ R(t) = \int_{-\infty}^{t} e^{A(t-s)}CX(s)ds + \int_{-\infty}^{t} e^{A(t-s)}BR(s)X(s)ds = \int_{-\infty}^{t} e^{A(t-s)}CX(s)ds + \int_{-\infty}^{t} e^{A(t-s)}BR(s)X(s)ds = \]
\[
\int e^{A(t-s)}CX(s)ds + \int e^{A(t-s)}B e^{A(s-\tau)}CX(r)X(s)drds + \int e^{A(t-s)}B e^{A(s-\tau)}BR(r)X(r)X(s)drds \tag{8}
\]
in turn. One can continue in this fashion to develop a series representation for the solution. The expansion will be convergent under appropriate conditions on A and B, see eg. Banks (1988), Chapter 4.

II. STEP-FUNCTION INPUT

In the case that the input is piecewise constant a convenient expression may be set down for the recursive computation of the state vector, S. Suppose that

\[ X(t) = X_n \quad \text{for } n-1 < t \leq n \]

and \( n = 0, \pm 1, \pm 2, \ldots \). Then the matricant of (4) is given by

\[ P(t,n) = \exp((t-n)(A + BX_n)) \]

for \( n < t \leq n+1 \). In terms of the state vector one has from (5)

\[ S(n) = \exp(A + BX_n)S(n-1) + (A + BX_n)^{-1}(\exp(A + BX_n) - I)a \tag{9} \]

This expression may be employed to compute the state vector in recursive fashion.

Next we will study the response of the system to pulses, to pulse-pairs and to piecewise constant noise. Some discussion about the relationships between parameters estimated by such experiments may be found in Blumberich and Kaiser (1983).

A. The Pulse Response

Consider the case where the input to the system is a unit pulse starting at time 0. Specifically let \( X(t) = 1 \) for \( 0 < t \leq 1 \) and equal 0 otherwise. Writing \( D = A + B \) one has from (9)

\[ S(1) = e^{D}S(0) + D^{-1}(e^{D} - I)a \]

and then

\[ S(t) = e^{A(t-1)}S(1) + A^{-1}(e^{A(t-1)} - I)a \tag{10} \]

for \( t \geq 1 \). In terms of \( R \), (10) becomes

\[ R(t) = e^{A(t-1)}E \tag{11} \]

where \( E = R(1) = (e^{D} - I)(D^{-1} - A^{-1})a \).

In the case that \( A \) is diagonal, with diagonal entries \( A_1, A_2 \) one has

\[ e^{At} = \begin{bmatrix} \exp(A_1t) & 0 \\ 0 & \exp(A_2t) \end{bmatrix} \]

Writing \( A_j = \beta_j + i\gamma_j \), the output, \( Y(t) = \Re(e^{\beta}S(t)) \) is seen to contain the frequencies \( \gamma_j \). It is apparent that these may be estimated by looking for peaks in the modulus of the (empirical) Fourier transform of output. The \( \beta_j \) might be estimated by nonlinear regression as in Bolt and Brillinger (1979). Expressions for the standard errors of the estimates are given in that paper.

B. The Two-Pulse Response

Consider next the case where the input to the system is

\[ X(t) = \begin{cases} 1 & \text{for } 0 < t \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

\[ u < t \leq u+1 \]

Let \( S(t,u) \) denote the system response as a function of \( t \) for given \( u \) and let \( R(t,u) = S(t,u) + A^{-1}a. \) Then from (11)
\[ R(t,u) = e^{A(t-u-1)} R(u+1) \]

while
\[ R(u+1) = e^D R(u) + E \]

and
\[ R(u) = e^{A(u-1)} E \]

giving
\[ R(t,u) = e^{A(t-u-1)} e^D e^{A(u-1)} E + e^{A(t-u-1)} E \]

In the case that \( D \) is not diagonal, the frequencies are coupled. This circumstance may be examined by taking the two-dimensional Fourier transform of the output \( Y(t,u) \). To be specific, consider the case of

\[ A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \] and \( B = \begin{bmatrix} B_1 & B_3 \\ 0 & B_2 \end{bmatrix} \]

Then \( D = A + B \) will also be upper triangular and so too will be
\[ e^D = \begin{bmatrix} \exp(D_1) & \delta \\ 0 & \exp(D_2) \end{bmatrix} \]

where \( \delta = B_3(\exp(D_1) - \exp(D_2))/(D_1 - D_2) \). Now \( Y(t+u,u) \) is
\[
-c^5 A^{-1} a + c_1 E_1 e^{D_1} e^{A_1(t-u-1)} e^{A_1(u-1)} + c_2 E_2 e^{D_2} e^{A_2(t-u-1)} e^{A_2(u-1)}
+ c_1 E_1 e^{A_1(t-1)} + c_2 E_2 e^{A_2(t-1)}
\]

Provided \( B_3 \) and so too \( \delta \neq 0 \) the output \( Y(t+u,u) \) may be anticipated to contain the (two-dimensional) frequency \((\gamma_1, \gamma_2)\). The absolute value of the two-dimensional Fourier transform will peak at that point.

### III. CROSS-CORRELATION ANALYSES

Next the case of a Gaussian input process will be considered. Suppose that one retains only the first two terms in (8). This would be reasonable if \( B \) or \( X \) were small. Then one has the approximate system response
\[ R(t) = \int g_1(r-s) X(s) ds + \int g_2(t-r, t-s) X(r) X(s) dr ds \]

with
\[ g_1(s) = e^{A_2} C \]

for \( s > 0 \) and
\[ g_2(r,s) = e^{A_2} B e^{A(r-s)} C \]

for \( r > s > 0 \). The corresponding Fourier transforms of \( g_1 \) and \( g_2 \) are
\[ G_1(\lambda) = (i\lambda I - A)^{-1} C \]

and
\[ G_2(\lambda, \mu) = (i (\lambda + \mu) I - A)^{-1} B (i\lambda I - A)^{-1} C \]

The first has peaks at \( \lambda = \gamma_j \) and the second at \( \lambda = \gamma_j, \lambda + \mu = \gamma_j \).

#### A. Linear Analysis
Following the cross-correlation analysis set down by Tick (1961), an estimate of the first-order transfer function corresponding to the impulse response \( h_1(t) = \text{Re}(e^{ig_1(t)}) \) may be derived as follows. Suppose that the input process, \( X \), is zero mean stationary Gaussian and that the error process \( \epsilon \) is uncorrelated with it. Write

\[
Y(t) = \int h_1(t-s)X(s)ds + \int\int h_2(t-r,t-s)X(r)X(s)drds + \epsilon(t)
\]

with \( h_2 \) symmetric in its two arguments. Then one has

\[
f_{yx}(\lambda) = H_1(\lambda)f_{xx}(\lambda)
\]

where \( f_{yx} \) and \( f_{xx} \) denote the cross-spectrum of the output with the input and the power spectrum of the input respectively.

B. Quadratic Analysis

Continuing to follow the development of Tick (1961) one has the following relationship, analogous to (17), for the quadratic transfer function, that is for the Fourier transform of \( h_2(u,v) = \text{Re}(e^{ig_2(u,v) + g_2(v,u)})/2 \),

\[
f_{xyy}(\lambda,\mu) = 2H_2(-\lambda,-\mu)f_{xx}(\lambda)f_{xx}(\mu)
\]

Here \( f_{xyy} \) is the cross-bispectrum of \( Y \) with \( X \). Details may be found in Brillinger (1970) and Feuerverger (1972). A procedure for estimating the cross-bispectrum is given in the Appendix.

Billmich and Zieessow (1983a,b) lay out the motivation for this type of analysis in the NMR case.

IV. MAXIMUM LIKELIHOOD ANALYSIS

The final estimation procedure to be considered is maximum likelihood. Suppose that the measurement equation (2) contains additive Gaussian white noise disturbance and so too does the transition equation. The likelihood of an observed stretch of data may be set down via the Kalman filter. It will be supposed that time is discrete \( n = 1, 2, 3, \ldots \) corresponding to piecewise constant input. The measurement and transition equations will be written

\[
Y_n = \text{Re}[e^{iS_n}] + \epsilon_n
\]

and

\[
S_n = A_nS_{n-1} + a_n + \xi_n
\]

respectively, where the processes \( \epsilon \) and \( \xi \) have variance \( \sigma^2 \) and variance-covariance matrix \( Q \) respectively. Here \( A_n = \exp(A + BX_n) \) and \( a_n = (A + BX_n)^{-1}(\exp(A + BX_n) - I)a \) following (9).

The log likelihood given data \( Y_1, ..., Y_n \) is

\[
\sum_{n=1}^{N} (v_n^2/b_n + \log b_n)
\]

with the quantities appearing able to be evaluated recursively as follows:

\[
v_n = Y_n - \text{Re}(e^{iS_n})
\]

\[
b_n = e^{P_n}c + \sigma^2
\]

\[
S_{n+1\mid n} = a_{n+1} + A_{n+1}(S_{n\mid n-1} + K_nv_n)
\]

\[
K_n = P_n\sigma/b_n
\]

\[
P_{n+1} = A_nP_nA_n^\top - A_nK_nK_n^\top A_n^\top b_n + Q
\]
The unknown parameters may be estimated by maximizing the likelihood.

In the case of NMR, Q may be taken to be 0 and the expressions simplify to

\[ v_n = Y_n - \text{Re}(e^{i\theta} S_{n+1|n-1}) \]
\[ b_n = \sigma^2 \]
\[ S_{n+1|n} = \lambda_{n+1} S_{n|n-1} + \eta_{n+1} \]
\[ K_n = 0 \]
\[ P_{n|n+1} = 0 \]

with \( S_{n|n-1} = S(n) \). The log likelihood is principally a sum of squares.

V. AN ELEMENTARY EXAMPLE

In order to better understand the relative merits of the various estimation procedures that have been described, a simple model has been set down and observations simulated. The parameter values employed are given in Table 1 below. The values were chosen to be somewhat similar to those of a particular NMR situation the author has been concerned with. The frequencies \( \gamma_1, \gamma_2 \) were take to be \( 2\pi/15 \) and \( 2\pi/5 \) radians per unit time respectively. These correspond to .067 and .2 cycles per unit time. The numbers of observations generated were multiples of 512. The process was started at the steady state value of \( -A^{-1}a \). Basic to the example is the fact that the frequencies \( \gamma_1 \) and \( \gamma_2 \) are coupled since \( b_3 \neq 0 \). It was further taken that \( e^2 = [1 \; 1] \) so that \( Y(t) = \text{Re}(S_1(t) + S_2(t)) + \epsilon(t) \). For the case of pulse input, the noise standard error, \( \sigma \), was taken to be .05 while for the stochastic input it was taken to be 3.0.

A. Results of the Linear Analysis

Figure 1 shows the impulse response and the absolute value of its Fourier transform based on 512 data points. Peaks are seen to occur near the frequencies .067 and .2 respectively, as was to be expected. The fact that the input and output series are discrete means that the peaks are broader than the ideal. The fact that the frequencies are coupled is not discernable on the basis of this plot.

Figure 2 presents the results of a stochastic simulation. The estimation procedure is that described in Section 3A and the Appendix. All told 50 periodograms based on data segments of length 512 were averaged. The top graph is the first segment of the output. The bottom graph gives \( |H^{-1}(\lambda)| \). There are again seen to be peaks near the frequencies .067 and .2 respectively.

B. Results of the Quadratic Analysis

Figure 3 presents a contour plot of the signal resulting from probing the system with pulses lag \( u \) apart, for various \( u \). The signal, rather than signal plus some noise, is graphed to make the behaviour clearer. The horizontal axis corresponds to time, while the vertical corresponds to lag, \( u \). The left-to-right slices give the responses starting at time 0. Values below the mean level have been contoured by dashed lines. It is apparent that after the initial pulse the system oscillates, as in Figure 2. Then when the second pulse arrives the output proceeds to oscillate in an altered fashion.

Figure 4 gives the absolute value of the two-dimensional Fourier transform of two-pulse output. Noise with \( \sigma = .05 \) was added. Results are displayed in both contour and perspective fashion. (Actually the final two terms of (12) were "removed" first to reduce leakage.) There are peaks near frequencies (.067,.067), (.2,.2) and (.067,.2). This last corresponds to coupling.

Next the case of stochastic input is considered. The data was the same as that of the linear analysis, except now 250 segments of length 512 were created and periodograms averaged. As an estimate of \( H_2(\lambda, \mu) \) one can consider \( f_{XX}(-\lambda, -\mu)/(2f_{XX}(\lambda)f_{XX}(\mu)) \).
Figure 5 displays $|H^{-2}(\lambda,\mu)|$, for the fundamental domain, in contour and perspective form. There is a peak at (-2,133) corresponding to the coupling introduced.

C. Results of the Maximum Likelihood Analysis

Next the results of a maximum likelihood analysis are presented. In setting up the unknown parameters it has been assumed that the experiments started with the system in steady state, i.e. $R(0) = 0$ and that the measurement set-up, i.e. $c$, is known. If these are unknown, their values may be included as further parameters to be estimated.

Only 512 observations were employed in this analysis and the estimates and their standard errors are given in Table 1 below. For some details of the computations see the Appendix.

Table 1

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>estimate</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>0.4189</td>
<td>0.4150</td>
<td>0.0022</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>1.2566</td>
<td>1.2525</td>
<td>0.0030</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.025</td>
<td>-0.02175</td>
<td>0.0024</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.004</td>
<td>-0.01066</td>
<td>0.00282</td>
</tr>
<tr>
<td>Re $a_1$</td>
<td>1.0</td>
<td>-0.987</td>
<td>0.7437</td>
</tr>
<tr>
<td>Im $a_1$</td>
<td>0.0</td>
<td>-0.4885</td>
<td>0.4594</td>
</tr>
<tr>
<td>Re $a_2$</td>
<td>1.0</td>
<td>1.544</td>
<td>0.5059</td>
</tr>
<tr>
<td>Im $a_2$</td>
<td>0.0</td>
<td>0.1290</td>
<td>0.4825</td>
</tr>
<tr>
<td>Re $b_1$</td>
<td>0.0</td>
<td>0.00950</td>
<td>0.00734</td>
</tr>
<tr>
<td>Im $b_1$</td>
<td>0.0</td>
<td>0.01391</td>
<td>0.00877</td>
</tr>
<tr>
<td>Re $b_2$</td>
<td>0.1</td>
<td>0.11005</td>
<td>0.01331</td>
</tr>
<tr>
<td>Im $b_2$</td>
<td>0.0</td>
<td>-0.00023</td>
<td>0.01201</td>
</tr>
<tr>
<td>Re $b_3$</td>
<td>1.0</td>
<td>0.6860</td>
<td>0.2129</td>
</tr>
<tr>
<td>Im $b_3$</td>
<td>0.0</td>
<td>0.2556</td>
<td>0.2141</td>
</tr>
<tr>
<td>$\sigma_3^2$</td>
<td>2.933</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results appear quite satisfactory. All the parameter values, except for $a_1$, have been fairly well-determined. The standard errors could have been driven lower by including more observations.

VI. DISCUSSION

Advantages of the one dimensional Fourier approach include: directness and availability of initial values for later more subtle analyses. Disadvantages are: peaks may be superposed and coupled frequencies cannot be discerned. Advantages of the two-dimensional analysis include: coupling may be discerned and other parameters may be estimated. Disadvantages are not all parameters of the system can be estimated and carrying out the data collection is complicated. Advantages of the full likelihood approach include: efficiency and available standard errors. An important disadvantage is that initial values for the optimization routine are required and these may be crucial.

Problems remaining include: design (choice of input $X$), development of conditions for consistency and asymptotic normality, and derivation of analytic expressions for standard errors.

ACKNOWLEDGEMENT

The author wishes to thank Professor Reinhold Kaiser for many helpful answers to questions concerning the NMR circumstance. The research was supported by NSF Grant MCS-8316634.
APPENDIX

A. Computation of the Spectrum Estimates

There are a variety of fashions in which spectra may be estimated. In the computations of this paper the technique of segmenting and averaging was employed. Specifically, let the data be broken into $L$ stretches of length $V$, so $T = LV$. Then compute the tapered Fourier transform of the $l$-th stretch,

$$ d^V_l(\lambda; l) = \sum_{v=0}^{V-1} \text{h}(\frac{v+1}{V+1}) Y(lV+v)e^{-iv\lambda} $$

for $l = 0,...,L-1$. (The taper employed in this work was the Hanning-Tukey, $h(u) = .54 - .46\cos(2\pi u)$.) In estimating the cross-spectrum, $f_{XY}(\lambda)$, the cross-periodogram of the $l$-th stretch is computed,

$$ I_{XY}^l(\lambda; l) = \frac{1}{2\pi V h_2} d^V_l(\lambda; l) d^V_l(\lambda; l) $$

where $h_2 = \sum h((v+1)/(V+1))^2$. The estimate of the cross-spectrum is now

$$ \hat{f}_{XY}(\lambda) = \frac{1}{L} \sum_{l=0}^{L-1} I_{XY}^l(\lambda; l) $$

To estimate the cross-bispectrum, $f_{XY}(\lambda, \mu)$, the following cross-biperiodogram is averaged

$$ I_{XY}^{\lambda\mu}(\lambda, \mu) = \frac{1}{(2\pi)^2 V h_3} d^V_l(\lambda; l) d^V_l(\mu; l) d^V_l(\lambda; l) $$

where $h_3 = \sum h((v+1)/(V+1))^3$.

B. The Likelihood Computations

The model simulated in the paper may be written

$$ Y_n = f_\star(\theta) + \varepsilon_n $$

with the $\varepsilon_n$ Gaussian white noise of mean 0 and variance $\sigma^2$. The parameter $\theta$ may be estimated by minimizing

$$ \sum_{n=1}^{N} (Y_n - f_\star(\theta))^2 $$

as a function of $\theta$. This is essentially minus the Gaussian log-likelihood. The covariance matrix of the estimate of $\theta$ may be estimated by the inverse of

$$ \sum_{n=1}^{N} \left[ \frac{df_\star(\hat{\theta})}{d\theta} \right] \left[ \frac{df_\star(\hat{\theta})}{d\theta} \right]^t $$

where $\hat{\theta}$ is the valued minimizing the sum of squares. Under regularity conditions the estimate will be asymptotically normal. The variance $\sigma^2$ may be estimated by dividing the minimized sum of squares by $N$.

The optimization routine employed in the computations reported is VA09A of the Harwell Subroutine Library.

REFERENCES


Figure Captions

Figure 1. The top graph gives the output resulting from unit pulse input to the system with parameters given in Table 1, except that the additive noise $\sigma$ is .05. The bottom graph gives the absolute value of the tapered Fourier transform of 512 points.

Figure 2. The top graph is the output resulting from standard Gaussian white noise input for the system with parameters given in Table 1. The bottom graph is the estimate of $|H_1(\lambda)|$ derived by averaging periodograms based on 50 segments of 512 points.

Figure 3. Two-pulse output, with the pulses lag $u$ apart. Values below the mean level are plotted via dashed curves.

Figure 4. The amplitude of the two-dimensional Fourier transform of the signal of Figure 3 with noise of standard error .05 added on. A taper was included in the Fourier transform.

Figure 5. The estimate of $|H_2(\lambda,\mu)|$ for the system with parameters in Table 1. The input was standard Gaussian white. Periodograms based on 250 segments of 512 points were averaged.
Pulse Response

Fourier Amplitude
Noise Response

Modulus Transfer Function
Two-Pulse Response
Two-Pulse Fourier Amplitude

![Graph showing Two-Pulse Fourier Amplitude with frequency on the x-axis and amplitude on the y-axis. The graph displays peaks at specific frequencies.]
Modulus Bitransfer Function

frequency

frequency
ACTAS DEL 2º CONGRESO LATINOAMERICANO DE PROBABILIDAD Y ESTADISTICA MATEMATICA

Sociedad Bernoulli - Comité Latinoamericano

Caracas - Julio de 1985
SOME STATISTICAL ASPECTS OF NMR SPECTROSCOPY

David R. Brillinger
Statistics Department, University of California
Berkeley, California, U.S.A.

ABSTRACT
Data collected in nuclear magnetic resonance (NMR) experiments is often Fourier transformed. Peaks in spectra so obtained are then used to identify and determine the structure of various complex molecules and sometimes their changes during reactions. The purpose of this paper is to describe the conceptual situation here in statistical terms and then to indicate an alternate method of analysis.

1. INTRODUCTION
Nuclear magnetic resonance (NMR) is a quantum mechanical phenomenon that may be exploited to study the structure of various compounds and substances. Briefly, when particular nuclei are placed in a strong magnetic field and then further perturbed by a radio-frequency field, a fluctuating voltage signal may be recorded. This signal reflects the absorption of energy by the nuclei present at characteristic resonance frequencies. In a pulse NMR experiment, the radio-frequency field is created by well-spaced pulses, the transients created (free induction decays) are averaged, and that average Fourier transformed. The Fourier transform values are examined for peaks corresponding to the resonance frequencies. In a two-pulse NMR experiment, the transients created by pulse-pairs are recorded as a function of time between pulses and time elapsed, (with perhaps a number of pulse-pair transients averaged). The two-dimensional Fourier transform of this function of two variables is computed and examined for peaks. Because the phenomenon involved is nonlinear, the result of a two-pulse experiment is not simply the sum of the single pulse results. Two-dimensional Fourier transform NMR spectroscopy is useful for examining the connectivity of coupled nuclei and thereby better determining molecular structure. A final type of NMR experiment to be mentioned is stochastic NMR spectroscopy. Here the radio-frequency field is randomly generated and cross-correlated with the measured signal. The correlation functions obtained (first and higher order) are then Fourier transformed and the resulting spectra examined for peak values as a function of frequency. Becker and Farrar (1972) and Levy and Craik (1981) are general references to the technique of NMR spectroscopy.
The two-pulse case is discussed in Aue et al. (1976), while stochastic NMR is reviewed in Blumich (1985).

The physics of the situation may be described as follows. The system development follows the von Neumann equation of motion for the density operator \( \rho \),

\[
\frac{d\rho(t)}{dt} = -i \left[ H(t), \rho(t) \right] + R(\rho(t) - \rho_0)
\]

[1]

see Bartholdi et al. (1976), Kaiser and Knight (1982). Here \( \rho(t) \) denotes the density operator, \( H(t) \) is the spin Hamiltonian (including any input stimulus), \( \rho_0 \) is the density operator at equilibrium, \( R \) is the relaxation operator and "[·]" signifies the commutator. The operator \( \rho(t) \) describes the state of the system at time \( t \). Equation [1] describes its evolution and this equation is seen to be bilinear in \( H(t) \) and \( \rho(t) \). If \( \mathbf{F} \) denotes the spin angular momentum operator, then the process observed may be denoted

\[
Y(t) = \text{tr} (\mathbf{F}\rho(t))
\]

[2]

with "\( \text{tr} \)" denoting trace. These various operators may be represented by matrices (of row or column dimension the number of basis states) and the commutator is \( [U, V] = 2\pi(U V - V U)/i\hbar, \) \( \hbar \) Planck's constant. A specific matrix formulation may be found in Grivet (1985). In the experiments of concern, the Hamiltonian takes the form \( H_0 + G X(t) \) with \( H_0 \) referring to the unstimulated system and \( X(t) \) a scalar stimulus process.

We reduce all the above discussion to the following system description and measurement equations,

\[
\frac{d\mathbf{S}(t)}{dt} = a + A\mathbf{S}(t) + B\mathbf{S}(t)X(t)
\]

[3]

\[
Y(t) = C\mathbf{S}(t)
\]

[4]

with \( \mathbf{S}(·) \) a state vector characterizing the system's evolution, with \( X(·) \) the system's input and \( Y(·) \) its output. In practice measurement noise will be present as well.

There is one particular case of the NMR phenomenon that has been much studied analytically and physically. This is the case of a two-level system with a phenomenological representation of relaxation.
The state vector, $S(t)$, of \([3]\) is three dimensional and the coefficients of equation \([3]\) have, in the rotating frame of reference, the form

$$
\dot{S} = \begin{bmatrix}
0 & -1/T_2 & \Delta \\
0 & -1/T_2 & 0 \\
-M_0/T_1 & 0 & -1/T_1
\end{bmatrix}
\begin{bmatrix}
S_1 \\
S_2 \\
S_3
\end{bmatrix}
$$

Here: $M_0$ denotes an equilibrium value, $T_1$ and $T_2$ are decay times and $\Delta$ denotes frequency offset. An elementary experimental setup has been assumed. Transverse magnetization is observed generally, referring to the first two coordinates of $S(t)$. In this case the equations \([3]\) are referred to as the Bloch equations.

2. **Solving the System Equation, \([3]\)**

In fact an exact expression may be set down for the solution of equation \([3]\), (but it must be remembered that solutions do not always exist, nor are they necessarily unique). Let $A(t) = A + Bx(t)$, then

$$
\frac{dS(t)}{dt} = \dot{\mathbf{a}} + A(t)S(t)
$$

Following Gantmacher (1959) the solution may be written

$$
S(t) = P(t,t_0)S(t_0) + \int_{t_0}^{t} P(t,u)du \mathbf{a}
$$

where $P(\cdot\cdot)$ denotes the matricant or state transition matrix. This last is the solution of

$$
\frac{dP(t,u)}{dt} = A(t)P(t,u), \quad P(u,u) = I
$$

In the case that $A(t)$ is constant, $-A\dot{P}(t,u) = \exp\{(t-u)A\}$. The matrix $P(\cdot\cdot)$ is generally complex, but has some simplifying properties. An expression may be determined for it iteratively in the equation \([8]\), namely

$$
P(t,u) = I = \int_{t}^{t} A(x)dx + \int_{u}^{t} x A(x)A(y)dy + \ldots
$$

This last is actually the multiplicative integral.
\[ P(t,u) = \int_t^u (I + A(x) dx) \]  

see Gantmacher (1959), Chapter XIV. The matrix \( P(\cdot) \) has the property

\[ P(t,u) = P(t,x)P(x,u) \]

hence, it follows that if \( A(\cdot) \) is piecewise constant

\[ A(t) = A_k \quad \text{for} \quad t_{k-1} < t \leq t_k \]

then

\[ P(t,t_0) = \exp\left\{ (t-t_{k-1})A_k \right\} \prod_{k=1}^{K-1} \exp\left\{ (t_k - t_{k-1})A_k \right\} \]

This expression is helpful in some circumstances.

We remark that Grunbaum (1985) sets down an expression for the solution of the Bloch equations, corresponding to (5) when \( T_1 = \infty \).

3. THE PULSE RESPONSE

Consider the case where the input to the system is the following pulse

\[ X(t) = \begin{cases} 1 & \text{for } 0 < t \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

Suppose \( t < 0 \) and that the system has been in operation long enough that a steady state has been reached. From [7]

\[ S(t) = \exp\left\{ (t-t_0)A \right\} + A^{-1}\left( \exp\left\{ (t-t_0)A \right\} - I \right)a \]

and the steady state must be \( S(t) = -A^{-1}a, \ t < 0. \) (This also follows directly from [6].) Suppose next that \( 0 < t \leq 1 \) and set \( D = A + B. \) Then from [7]

\[ S(t) = \exp\{tD\}S(0) + D^{-1}\left( \exp\{tD\} - I \right)a \]

and oscillations at frequencies corresponding to the imaginary parts of the eigenvalues of \( D \) will occur. Suppose, finally, that \( 1 < t. \) From [7]

\[ S(t) = \exp\{(t-1)A\}S(1) + A^{-1}\left( \exp\{(t-1)A\} - I \right)a \]

This is the desired single pulse response.
Some practical implications of this last are the following: in the case that the eigenvalues, \( \lambda_j \), of \( A \) are distinct one has

\[
\exp(tA) = \sum_j Z_j \exp(\lambda_j t)
\]

where the \( Z_j \)'s are polynomials in \( A \) (see Gantmacher (1959), Chapter V), for stability one will want \( \Re(\lambda_j) < 0 \) and assuming measurement noise may be present, the output will have the form

\[
Y(t) = \sum_j a_j \exp(-\beta_j t) \cos(\gamma_j t + \delta_j) + \epsilon(t), \quad t > 1
\]

having written \( \lambda_j = -\beta_j + i\gamma_j \), and \( \epsilon \) denoting noise. The empirical estimation of the parameters of the model [18] is studied in Bolt and Brillinger (1979). If a number of pulse responses are averaged, the model [18] will retain its form, but the noise may be expected to be reduced.

For the Bloch equations, corresponding to [5] the eigenvalues are \(-1/T_1 \) and \(-1/T_2 + i\Delta \).

The difficulty that can arise in practice with this approach is that \( \gamma_j \) are too close and \( \beta_j \) too large for effective resolution.

4. THE TWO PULSE RESPONSE

Consider next the case where the input to the system is

\[
X(t) = 1 \quad \text{for } 0 < t < 1 \quad \text{and} \quad u < t < u+1
\]

\[
= 0 \quad \text{otherwise}
\]

Suppose \( t \geq u+1 \). From expression [7]

\[
\tilde{S}(t) = \exp((t-u-1)A)\tilde{S}(u+1) + A^{-1}(\exp((t-u-1)A) - I)\tilde{S}(1)
\]

while

\[
\tilde{S}(u+1) = \exp(D) \tilde{S}(u) + D^{-1}(\exp(D) - I)\tilde{S}(1)
\]

and

\[
\tilde{S}(u) = \exp((u-1)A) \tilde{S}(1) + A^{-1}(\exp((u-1)A) - I)\tilde{S}(1)
\]

Putting expressions [20] - [22] together and writing \( t-u = \nu \) leads to the following form for the output
\[ Y(u,v) = \sum_{j} \sum_{j'} a_{jj'} \exp(-B_j u - B_{j'} v) \cos(\gamma_j u + \gamma_{j'} v + \delta_{jj'}) \]

plus similar terms in \( \gamma_j \), \( \gamma'_{j'} \), plus a noise term. The amplitude of the empirical Fourier transform of such a \( Y(u,v) \) may be expected to be large near the points \((\gamma_j, \gamma'_{j'}), (\gamma_j, 0), (0, \gamma'_{j'})\) and so may be employed to estimate those values.

It appears that the estimation procedure of Bolt and Brillinger (1979) extends readily to this higher dimensional situation and may be employed to estimate the standard errors of the principal estimates.

5. STOCHASTIC INPUT

Ernst (1970) and Kaiser (1970) introduced the technique of stochastic NMR spectroscopy. The input is a random process selected by the experimenter. Resonance frequencies are determined by first cross-correlating the input and outputs of the system and then Fourier transforming the result. The system equation here is

\[
\frac{dS(t)}{dt} = a + AS(t) + BS(t)X(t) \tag{23}
\]

and the measurement equation \( Y(t) = CS(t) \). It is clear that the cross-correlation of the input and output may be determined from the cross-correlation of \( X(\cdot) \) and \( S(\cdot) \). We shall study the case in which \( X(\cdot) \) is piecewise constant

\[ X(t) = X_n \quad \text{for } n-1 < t \leq n \tag{24} \]

with \( X_n \) taken to be (statistically) independent of \( S(u) \) for \( u \leq n-1 \) and of \( X_m \) for \( m < n \). The \( X_n \) will be further assumed identically distributed, zero mean, and finally it will be assumed that the process \( Y(\cdot) \) is observed only at the times \( t = 0, 1, 2, \ldots \). One has the representation

\[ S(n) = \exp(A + BX_n) S(n-1) + (A + BX_n)^{-1}(\exp(A + BX_n) - I)a \tag{25} \]

Under the assumed conditions one sees from \(25\) that \( \text{cov}(S(n), X_{n-1}) = \emptyset \text{ cov}(S(n-1), X_{n-1}) \), \( \text{cov}(S(n), X_{n-2}) = \emptyset \text{ cov}(S(n-1), X_{n-2}) \), \( \emptyset \text{ cov}(S(n-2), X_{n-2}) \) and so on, with
\[ \Phi = \mathbb{E}(\exp(A + BX)) \]  

In general one has

\[ \text{cov}(S(n), X_m) = \Phi^{n-m} \text{cov}(S(n), X_n) \quad m < n \]  

The (discrete) Fourier transform of this is proportional to \( (e^{i\lambda} - \Phi)^{-1} \), with \( \lambda \) denoting frequency. It may be expected to have peaks for \( \lambda \) near the imaginary parts of the eigenvalues of \( \log I \).

In the case that \( X_n \) are independent normals with variance \( \sigma^2 \) direct evaluation yields

\[ \Phi = \exp(A + \sigma^2 B^2/2) \]  

in accord with the result of Kaiser and Knight (1982), although the set-ups are somewhat different. The sample Fourier transform will be evidencing the imaginary parts of the eigenvalues of \( A + \sigma^2 B^2/2 \). If \( X_n = iX \) each with probability 1/2,

\[ \Phi = (\exp(A + BX) + \exp(A - BX))/2 \]  

whose eigenvalues relate rather indirectly, in general, to those of \( A \).

It is to be noted that Kaiser and Knight (1982) also develop an expression for the higher-order cross moments in the case of white normal input. In general such moments might best be developed via a Volterra expansion calculus, as in Brillinger (1970).

It is further worth remarking that the measurement equation [4] can become

\[ Y(t) = CS(t) + \varepsilon(t) \]  

\( \varepsilon(t) \) an external noise, with no change in the cross-correlation function.

6. MAXIMUM LIKELIHOOD ESTIMATION

The techniques of the preceding sections have concentrated on the estimation of just some of the characteristics of the matrices \( a, A, B \). In this section a maximum likelihood procedure is indicated that is directed at estimating \( a, A, B \) completely (as characterized totally by some collection of parameters). The approach has the further advantage of allowing both the system and measurement equations to include noise.
terms and of producing an estimate of the state vector itself. The procedure is based upon the Kalman filter.

Suppose again that the input is piecewise constant, as in (24). Let the equation (25) have a noise term and be written

\[ S_n = A_n S_{n-1} + a_n + \xi_n \]  
\[ Y_n = CS_n + \epsilon_n \]

with \( A_n = \exp(A + BX_n) \), with \( a_n = (A + BX_n)^{-1}(\exp(A + BX_n) - I)a \), and

with \( \xi_n \) the noise term. Let the measurement equation be written

\[ S_n = A S_{n-1} + a_n + \xi_n \]  
\[ Y_n = CS_n + \epsilon_n \]

\( n = 1, 2, \ldots, N \). Assume that the series \( \xi_n \) and \( \epsilon_n \) are independent Gaussian white noises with \( \text{var} \, \xi_n = Q \) and \( \text{var} \, \epsilon_n = r \). Let the various unknowns in \( a, A, B, Q, r \) be denoted by \( \Theta \). Then \( \Theta \) may be estimated by maximizing the likelihood, which may be evaluated (for any value of \( \Theta \)) as follows.

Minus twice the log likelihood is given by

\[ N \sum_{n=1}^{N} \left( \frac{v_n^2}{b_n} + \log b_n \right) \]  
\[ v_n = Y_n - CS_n \]
\[ b_n = CP_nC^T + r \]

The \( v_n, b_n \) may be evaluated by the following recursions, which introduce some supplementary variables:

\[ S_{n+1|n} = a_{n+1} + A_{n+1} (S_{n|n-1} + K_n v_n) / b_n \]  
\[ K_n = P_nC^T \]
\[ P_{n+1} = A_n P_n A^T - A_n K_n K_n^T A^T / b_n - Q \]

The equations (34) are known as the Kalman filter. The Kalman filter is discussed in detail in Kailath (1981). The associated maximum likelihood approach was introduced in Mehra (1971). Various statistical aspects are discussed in Hannan (1979). An incidental output of the procedure is \( S_n|n = S_{n|n-1} + K_n v_n / b_n \), an estimate of the state vector at time \( n \).

Further details of this approach will be provided in another paper.
An essential difference from the various spectral approaches is that all the parameters of the system are being estimated and they are being estimated by a method that has been much studied, maximum likelihood. In some circumstances the maximum likelihood computations may be carried through by the EM algorithm, see Shumway and Stoffer (1982).

ACKNOWLEDGMENT
This work was supported by the National Science Foundation Grant DMS-8316634.

REFERENCES

Science 178, 361 (1972).


[4] Blumich, B.


[8] Gantmacher, F.R.


[12] Kaiser, R.

J. Spund Vib. 12, 301 (1970).


[16] Hannan, E.J.

[17] Shumway, R.H. & Stoffer, D.S.
TECHNICAL REPORTS
Statistics Department
University of California, Berkeley


32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.


43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?


56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.


64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.


71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.


77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.


90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.


95. CANCELLED


114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.
118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.
127. Same as No. 133.
128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.
131. Same as No. 140
132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.