MODELING NONSTATIONARY RANDOM PROCESSES WITH AN APPLICATION TO GYRO DRIFT RATE

by

Albert John Van Dierendonck

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Electrical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Ames, Iowa

1968

TABLE OF CONTENTS

I. INTRODUCTION

During the last fifteen years, many of the techniques developed to handle stationary random processes in filter and control theory have been extended or developed to handle nonstationary random processes (3, 6, 7, 10, 17, 18). The "Kalman" filtering technique developed more recently will also handle nonstationary processes (26).

In both cases, a priori statistics are required. In the first case, the a priori statistics are the covariance functions and the mean value functions describing the processes. In the case of "Kalman" filtering, the a priori statistics are in the form of transition matrices and input covariance matrices describing the processes (26).

The covariance and mean value functions are expectations of a function of the random processes. If these processes are ergodic covariance stationary random processes, these expectations can be estimated with time averages (24).

The transition and input covariance matrices can be computed from these expectations (26). For example if a process has a rational power spectral density function, this power spectral density function can be factored to determine the filter that a "white noise" process is passed through to obtain this process. The transition matrix and input covariance matrix describing this process can then be computed from the transfer function of the filter.

Conversely, the covariance function can be computed from the transition and input matrices. This is shown in Chapter III of this dissertation.

Due to the frequent lack of a large ensemble of sample functions of a process, and due to the computational difficulties in calculating ensemble averages, the need for other techniques is apparent.

The purpose of this study is to present a technique for estimating a priori parameters of a class of random processes from experimental data in such a way that time averages can be used. This class of random processes is one which can be thought of as passing stationary "white noise" through a time-invariant linear filter. This class includes those nonstationary processes which are outputs of unstable time-invariant linear filters, or which are transient outputs of stable filters, as well as stationary processes, which are steady-state outputs of stable filters.

Hypothesized examples of these random processes are gyro drift rate (8, 14, 15), which motivated this study, and displacement of a particle under Brownian motion (27). Another example is a process which is a "response of linear networks to suddenly applied stationary random noise" (18).

II. DEFINITIONS

A. Random Variables

1. Definition of a random variable

A random variable X is a real finite valued function defined on a sample description space S with a probability function $P_X[\cdot]$ defined on a family $\mathcal{F}$ of events to which S belongs, where for every Borel set B of real numbers the set $\{s:X(s) \text{ is in } B\}$ is an event in $\mathcal{F}$.

The probability function of X, $P_X[\cdot]$, defined on every Borel set B of real numbers is the probability that an observed value of X is in B (24).

Let $\underline{X}$ be a k dimensional random variable. Then it may be characterized by its joint distribution function

$$F_{\underline{X}}(\underline{x}) = P_{\underline{X}}[X_i \leq x_i;\ i = 1, \cdots, k] \tag{2.1}$$

where $\underline{x} \in R^k$ (k dimensional real space), and $P_{\underline{X}}[\cdot]$ is the probability function of $\underline{X}$ (23).

2. Definition of the expectation of a random variable

If $\underline{Y} = \varphi(\underline{X})$ is an n dimensional random variable and a function of $\underline{X}$ (a k dimensional random variable), then

$$E[\varphi(\underline{X})] = \int \cdots \int_{R^k} \varphi(\underline{x})\ dF_{\underline{X}}(\underline{x}) \tag{2.2}$$

is the expectation of $\underline{Y} = \varphi(\underline{X})$, if it exists (23).

B. Random Processes

1. Definition of a random process

A random process is defined to be a family of random variables $\{X(t),\ t\in T\}$ indexed by a parameter t varying in an index set T (24).

a. Mean value function Let $\{\underline{X}(t),\ t\in T\}$ be a k-dimensional random process. The mean value function of $\underline{X}(t)$ is also k-dimensional, where

$$\underline{m}_{\underline{X}}(t) = E[\underline{X}(t)] \tag{2.3}$$

is defined for all t in T, provided that the expectation exists.

b. Covariance function Let $\{\underline{X}(t),\ t\in T\}$ be a column k-vector random process. The covariance function of $\underline{X}(t)$ is a kxk matrix defined by

$$\begin{aligned} K_{\underline{X}}(s,\ t) &= E\{[\underline{X}(t) - \underline{m}_{\underline{X}}(t)][\underline{X}(s) - \underline{m}_{\underline{X}}(s)]'\} \\ &= \mathrm{cov}[\underline{X}(s),\ \underline{X}(t)] \end{aligned} \tag{2.4}$$

where (′) indicates transpose, and $K_{\underline{X}}(s,\ t)$ is defined for all s, t in T, provided that the expectation exists. If $\{X(t),\ t\in T\}$ is a scalar random process, its variance function is then

$$\sigma_X^2(t) = K_X(t,\ t) \tag{2.5}$$

c. Ensemble average of a random process Let $\{X(t),\ t\in T\}$ be a random process. Then $\{X(t),\ t\in T\}$ may be written as the collection of sample functions $X_i(t)$, $t\in T$, $i = 1, \cdots, n$. Thus, if $Y_i = \varphi(X_i(t))$, $i = 1, \cdots, n$, is a function of $X_i(t)$, $i = 1, \cdots, n$, the ensemble average of the random process $\{Y(t),\ t\in T\}$ is given as (5)

$$\langle Y(t)\rangle = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Y_i(t)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i(t)), \quad t\epsilon T \tag{2.6}$$

C. Linear Dynamical Systems

1. Definitions (16)

a. Discrete-time linear dynamical systems A discrete-time linear dynamical system D is the system of difference equations

$$\text{D)} \quad \underline{x}(t) = A(t, t-1)\,\underline{x}(t-1) + B(t, t-1)\,\underline{u}(t)$$

$$\underline{y}(t) = H(t)\,\underline{x}(t), \quad t\epsilon T \tag{2.7}$$

where $\underline{x}\epsilon R^k$ is the state of the system, $\underline{u}\epsilon R^m$ is the input to the system, $\underline{y}\epsilon R^n$ is the output of the system, H is the output matrix of the system, and the index set T is the set of positive integers.

1. Transition matrix A of Equation 2.7 is the transition matrix of system D.

2. Input matrix B of Equation 2.7 is the input matrix of system D.

If A, B, and H are constant matrices, system D is said to be time-invariant.

b. Continuous-time linear dynamical systems A continuous-time linear dynamical system L is the system of differential equations

$$\text{L)} \quad \dot{\underline{x}}(t) = F(t)\,\underline{x}(t) + G(t)\,\underline{u}(t)$$

$$\underline{y}(t) = H(t)\,\underline{x}(t), \quad t\epsilon\tilde{T} \tag{2.8}$$

where $\underline{x} \epsilon R^k$ is the state of the system, $\underline{u} \epsilon R^m$ is the input to the system, $\underline{y} \epsilon R^n$ is the output of the system, H is the output matrix of the system, and the index set $\tilde{T}$ is the set of positive real numbers.

1. Stability matrix F of Equation 2.8 is the stability matrix of system L.

2. Input matrix G of Equation 2.8 is the input matrix of system L.

If F, G, and H are constant matrices, system L is said to be time-invariant.

The general solution of Equation 2.8 is

$$\underline{x}(t) = \Phi(t, t_o)\, \underline{x}(t_o) + \int_{t_o}^{t} \Phi(t, \tau)\, G(\tau)\, \underline{u}(\tau) d\tau, \quad t \epsilon \tilde{T} \tag{2.9}$$

where, if L is time-invariant,

$$\Phi(t, t_o) = \Phi(t - t_o) = e^{F[t - t_o]}, \quad t \epsilon \tilde{T} \tag{2.10}$$

The exponential function of a matrix F is defined to be

$$e^{F[t - t_o]} = \sum_{i=o}^{\infty} \{F[t - t_o]\}^i / i!, \quad t \epsilon \tilde{T} \tag{2.11}$$

where $\{F[t - t_o]\}^o = I$.

3. Transition matrix Φ of Equation 2.9 is the transition matrix of system L.

A time-invariant transition matrix may also be found from Laplace transforms by

$$\Phi(t) = \mathcal{L}^{-1}\{[sI - F]^{-1}\} \tag{2.12}$$

where $\mathcal{L}^{-1}$ denotes inverse Laplace transform (12).

III. THEORETICAL INVESTIGATION

A. Formulation of a Model

The type of model used to describe a random process depends on the application which motivated the modeling. In this case, the application of primary interest is one where gyro drift rate is part of an overall process which is to be estimated by means of a "Kalman" filter.

The "Kalman" filter utilizes Equation 2.7 or Equations 2.8 and 2.9 in a recursive scheme to weight in an optimum way past information along with a present noisy measurement to estimate the state of the system. In this scheme we utilize the vector difference equation

$$\underline{x}_{n+1} = \Phi_n \underline{x}_n + \underline{g}_n, \quad n = 0, 1, 2 \cdots \tag{3.1}$$

where Φ_n is a transition matrix and $\underline{g}_n$ is an input vector, which is the response due to a "white noise" driving function that is independent of $\underline{x}_n$.

If the system is a discrete-time system,

$$\underline{g}_n = B(n+1, n)\, \underline{u}(n), \quad n = 0, 1, 2, \cdots \tag{3.2}$$

where $\{\underline{u}(t), t \epsilon T\}$ is an independent random sequence. If the system is a continuous-time system,

$$\underline{g}_n = \int_{t_n}^{t_{n+1}} \Phi(t_{n+1}, \tau) G(\tau)\, \underline{u}(\tau) d\tau, \quad n = 0, 1, 2, \cdots \tag{3.3}$$

where $\{\underline{u}(t), t \epsilon \tilde{T}\}$ is a "white noise" random process. In either case,

$$K_n = E[\underline{g}_n \underline{g}_n'], \quad n = 0, 1, 2, \cdots \tag{3.4}$$

is the input covariance matrix. From this we can see that the input covariance matrix can be found if the input matrix and transition

matrix are known.

In this case, the gyro drift rate is an input. It is not a "white noise" process nor is it an independent random sequence. However, if we can express this correlated process in the form of Equation 3.1, we can augment the total system such that all inputs to it are either "white noise" processes or independent random sequences (26).

The original approach to the problem of modeling a nonstationary process was to somehow model it using time averages, and then worry about transforming whatever model we have to the form of Equation 3.1. Due to fact that experimental data are rarely in continuous form, and that computer solutions force us to consider a discrete form, the model that proved to be very susceptible to the use of time averages is that of a constant coefficient k^{th} ordered difference equation with an ergodic covariance stationary random input sequence. That is, the susceptible model is of the form

$$a_o X(t) + a_1 X(t - 1) + \cdots + a_k X(t - k) = N(t), \quad t \epsilon T \tag{3.5}$$

where T is the set of positive integers, $\{X(t), t \epsilon T\}$ is the discrete process being investigated, and $\{N(t), t \epsilon T\}$ is the input sequence. The time averaging is discussed in the next section.

We can rewrite Equation 3.5 in the matrix form

$$\underline{X}(t) = A\underline{X}(t - 1) + \underline{b}N(t), \quad t \epsilon T \tag{3.6}$$

where

$$\underline{X}(t) = \begin{bmatrix} X(t - k + 1) \\ \vdots \\ X(t - 1) \\ X(t) \end{bmatrix}, \quad t \epsilon T \tag{3.7}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & & 1 \\ a_{k1} & a_{k2} & & & a_{kk} \end{bmatrix} \tag{3.8}$$

and

$$\underline{b} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ b_k \end{bmatrix} \tag{3.9}$$

The entries of A and $\underline{b}$ are

$$a_{k(k-i+1)} = -\frac{a_i}{a_o}$$

and (3.10)

$$b_k = \frac{1}{a_o}$$

If $\{N(t),\ t\epsilon T\}$ is an independent random sequence, Equation 3.6 is of the form of Equation 3.1, where

$$\underline{x}_{n+1} = \underline{X}(t)$$

$$\Phi_n = A \tag{3.11}$$

$$\underline{g}_n = \underline{b}N(t)$$

and

$$t = n + 1,\ n = 0,\ 1,\ 2,\ \cdots$$

Furthermore, it can be shown that A has the necessary properties of a

transition matrix for integer values of t and n. In particular

$$A(t, t - \alpha) = A(\alpha) = [A(1)]^{\alpha} = [A(t, t - 1)]^{\alpha} \tag{3.12}$$

where α is a positive integer and $A(o) = A(t, t) = I$.

B. Determination of the Model

The model to be determined for a sequence $\{X(t), t\epsilon T\}$ is that of Equation 3.5. This is done by finding the order k, the constants $(a_o, \cdots, a_k)$, and the sequence $\{N(t), t\epsilon T\}$, which best describe the random sequence $\{X(t), t\epsilon T\}$. When dealing with experimental data we, of course, do not know how well this model will fit the process at hand; so we will simply try it, empirically, and see what results.

If we knew the input sequence $\{N(t), t\epsilon T\}$, this would be a problem of system identification. However, in this case, $\{N(t), t\epsilon T\}$ is not known, although its covariance function is hypothesized to be

$$\begin{aligned} K_N(s, t) &= \sigma^2 \delta(s, t) \\ &= \sigma^2, \; s = t \\ &= 0, \; s \neq t \end{aligned} \tag{3.13}$$

for all $t, s\epsilon T$, where $\delta(s, t)$ is the Kronecker delta. Furthermore,

$$\begin{aligned} K_N(s, t) &= K_N(s - t) \\ &= K_N(v) \\ &= \sigma^2 \delta(v, o), \; v\epsilon T \end{aligned} \tag{3.14}$$

where $v = s - t$.

Thus, since $\{N(t), t\epsilon T\}$ has these properties, so does the left hand side of Equation 3.5. First of all, from Equation 3.14, we have

$$K_N(v) = \operatorname{cov}\Big[\sum_{i=o}^{k} a_i X(t-i),\ \sum_{i=o}^{k} a_i X(t+v-i)\Big]$$

$$= E\Big[\sum_{i=o}^{k}\sum_{j=o}^{k} a_i a_j X(t-i)\, X(t+v-j)\Big],\ v\epsilon T \qquad (3.15)$$

But,

$$E\Big[\sum_{i=o}^{k}\sum_{j=o}^{k} a_i a_j X(t-i)\, X(t+v-j)\Big]$$

$$= \sum_{i=o}^{k}\sum_{j=o}^{k} a_i a_j E[X(t-i)\, X(t+v-j)],\ v\epsilon T \qquad (3.16)$$

In general, $\{X(t),\ t\epsilon T\}$ is a nonstationary random sequence. Thus, the expectation in Equation 3.16 must be an ensemble average (5). As stated in Chapter I, we wish to avoid estimating the statistics of a random process by ensemble means.

However, because of the ergodicity of $\{N(t),\ t\epsilon T\}$, the covariance function of Equation 3.15 may be estimated by a finite-time average using just one sample function of $\{N(t),\ t\epsilon T\}$ (24). That is,

$$K_N(v) \approx K_{T_1N}(v) = \frac{1}{T_1}\sum_{t\epsilon T_1} N(t)\, N(t+v)$$

$$= \frac{1}{T_1}\sum_{t\epsilon T_1}\sum_{i=o}^{k}\sum_{j=o}^{k} a_i a_j X(t-i)\, X(t+v-j),\quad v\epsilon T \qquad (3.17)$$

for large T_1, where T_1 is the range of summation on t.

If the a_i's and the order k in Equation 3.5 were known exactly, and if a Gaussian sequence is assumed, then $K_{T_1N}(v)$ in Equation 3.17 would converge in the mean to $K_N(v)$ as T_1 approaches infinity. For T_1 sufficiently large, the variance of the error for each v is on the

order of $\frac{1}{T_1}$ (19).

But the a_i's and order k in Equation 3.5 are not known. However, if we choose an order k and constants $(a_o, \cdots, a_k)$, we can express the error of this choice, plus the error due to the finite-time averaging, as a function of v by

$$E_{T_1}(v) = K_N(v) - \frac{1}{T_1} \sum_{t \epsilon T_1} \sum_{i=o}^{k} \sum_{j=o}^{k} a_i a_j X(t - i)\, X(t + v_j), \quad v \epsilon T$$

$$= K_N(v) - \hat{K}_{T_1 N}(v) \tag{3.18}$$

The estimation error due to the finite-time averaging is fixed for a fixed T_1. Thus, it is a lower bound for $E_{T_1}(v)$, $v \epsilon T$. The problem is to minimize the error $E_{T_1}(v)$ for all v with the proper choice of the a_i's and the order k in Equation 3.5. This can be accomplished by a gradient technique on an appropriate error function. This will be discussed in Chapter IV.

C. Criteria

The error for each v in Equation 3.18 is obviously non-increasing with each higher order k, when minimized by the proper choice of $(a_o, \cdots, a_k)$. Thus, a criterion is necessary to select a satisfactory order k. An excellent criterion is to choose a model of order k that results in a satisfactory estimation of $K_N(v)$, $v \epsilon T$, of Equation 3.14, with the proper choice of $(a_o, \cdots, a_k)$.

This criterion will vary with different application. For instance, if the sampling rates, time constants, etc. of the system application are such that the covariances at the certain covariance times $v \neq o$ are

not critical to system operation, the covariance estimations $\hat{K}_{T_1N}(v)$ at these times may be significantly greater than zero, and the criterion may be relaxed.

Otherwise, the goodness of $\hat{K}_{T_1N}(v)$, $v \neq o$, can be measured as follows. Suppose the sequence $\{N(t), t\epsilon T\}$ was truly a sequence of Gaussian random variables with zero mean and a covariance function described by Equation 3.14. Then the estimated correlation function

$$\hat{\rho}(v) \equiv \frac{\hat{K}_{T_1N}(v)}{\hat{K}_{T_1N}(o)}, \quad v\epsilon T \tag{3.19}$$

may be transformed to the function

$$T(v) = \hat{\rho}(v)\left(\frac{T_1 - 2}{1 - [\hat{\rho}(v)]^2}\right)^{\frac{1}{2}}, \quad v\epsilon T \tag{3.20}$$

which has a "Student's" t distribution with $T_1 - 2$ degrees of freedom (22). Therefore, for large T_1, the probability

$$P_\epsilon = P[|\hat{\rho}(v)| > \epsilon] = P\left[|T(v)| > \epsilon\left(\frac{T_1 - 2}{1 - \epsilon^2}\right)^{\frac{1}{2}}\right]$$

$$= 2 - 2\Omega\left[\epsilon\left(\frac{T_1 - 2}{1 - \epsilon^2}\right)^{\frac{1}{2}}\right], \quad v\epsilon T \tag{3.21}$$

is the probability that the absolute value of $\hat{\rho}(v)$ is greater than $\epsilon > o$, where $\Omega(x)$ is the distribution function of a standard normal random variable.

Equation 3.21 will give us an idea of how good our correlations would be if our model were perfect, thus, in a sense, supplying us with a measure of goodness. For example, if we choose an ϵ that results in a

negligible P_ϵ for the sample size T_1 used, we should make sure that a high percentage of the applicable correlations are within the bounds $|\hat{\rho}(v)| \leq \epsilon$, and that the correlations not within these bounds are not much greater than ϵ.

D. Augmentation for "Kalman" Filtering

1. Definition of the augmented system

If a random sequence $\{X(t),\ t\epsilon T\}$ is described by Equations 3.7 through 3.10, where this sequence is the input to a system, this system can be augmented such that $\underline{X}(t)$ becomes a state vector of the system, and all of the system inputs are "white noise" sequences (26). That is, the system is augmented to the form

$$\begin{bmatrix} \underline{Y}(\xi) \\ \underline{X}(\xi) \end{bmatrix} = \begin{bmatrix} \Phi(\xi,\ \xi\text{-}1) & \Lambda(\xi,\ \xi\text{-}1) \\ 0 & \Theta(\xi,\ \xi\text{-}1) \end{bmatrix} \begin{bmatrix} \underline{Y}(\xi\text{-}1) \\ \underline{X}(\xi\text{-}1) \end{bmatrix} + \begin{bmatrix} \underline{Z}(\xi,\ \xi\text{-}1) \\ \underline{W}(\xi,\ \xi\text{-}1) \end{bmatrix},\ \xi\epsilon T \tag{3.22}$$

where, in the ξ domain, the m-vector

$$\underline{Y}(\xi) = \Phi(\xi,\ \xi - 1)\ \underline{Y}(\xi - 1) + \Lambda(\xi,\ \xi - 1)\ \underline{X}(\xi - 1) + Z(\xi,\ \xi - 1)$$

and (3.23)

$$\underline{X}(\xi) = \Theta(\xi,\ \xi - 1)\ \underline{X}(\xi - 1) + \underline{W}(\xi,\ \xi - 1),\ \xi\epsilon T$$

$\underline{Z}(\xi,\ \xi - 1)$ is an m-vector of other "white noise" input sequences.

The problem with this form is that the sequence in Equations 3.7 through 3.10 is defined for integer values of t only. This is, of course, normalized from the sampling rate at which the sample sequence X(t) was measured. If the scale factor d_ξ of ξ is an integer (α) multiple of

the scale factor d_t of t (normalized from the sampling rate), then

$$\Theta(\xi, \xi - 1) = A^{\alpha}, \quad \xi \epsilon T \tag{3.24}$$

and

$$\underline{W}(\xi, \xi - 1) = \underline{M}(\xi), \quad \xi \epsilon T \tag{3.25}$$

is an independent random vector sequence with a covariance matrix

$$K_{\underline{M}} = \sigma^2 \sum_{j=o}^{\alpha-1} A^j \underline{b}\, \underline{b}'\, A'^j \tag{3.26}$$

which is the input covariance matrix of the vector difference equation

$$\underline{X}(\xi) = A^{\alpha} \underline{X}(\xi - 1) + \sum_{j=o}^{\alpha-1} A^j \underline{b} N(\xi - j), \quad \xi \epsilon T \tag{3.27}$$

which is the same as

$$\underline{X}(t) = A^{\alpha} \underline{X}(t - \alpha) + \sum_{j=o}^{\alpha-1} A^j \underline{b} N(t - j), \quad t = \alpha, 2\alpha, \cdots \tag{3.28}$$

The coupling matrix which couples $X(\xi)$ into the system is

$$\Lambda(\xi, \xi - 1) = \begin{bmatrix} 0 \cdots 0 & \gamma_{1k} \\ \vdots & \vdots \\ 0 \cdots 0 & \gamma_{mk} \end{bmatrix}, \quad \xi \epsilon T \tag{3.29}$$

with one γ_{ik} not equal to zero.

However, if the scale of ξ is not an integer multiple of the scale of t, $\Theta(\xi, \xi - 1)$, $\underline{W}(\xi, \xi - 1)$ and $\Lambda(\xi, \xi - 1)$ are not of this form. Therefore, they must be defined in another way. This is discussed in detail in Appendix C.

E. Covariance Functions

If the difference equation model has been determined, we have, by using the method described in Section D of this chapter, the model of $\{X(t), t\epsilon T\}$ in a form necessary to be used in the "Kalman" filtering technique. It is not, however, of the form used in the filtering techniques of Wiener (5), Boonton (3), Darlington (6), Koschmann (17), Davis (7), Lampard (18), Friedland (10), etc., where the covariance function of the sequence $\{X(t), t\epsilon T\}$ is utilized.

1. Numerical determination

If only the numerical values of the covariance function $K_X(s, t)$ of the random sequence $\{X(t), t\epsilon T\}$ are necessary for each $s, t\epsilon T$, we can determine this covariance function directly from Equation 3.6. This equation can be recursively reduced to the form

$$\underline{X}(t) = A^t\underline{X}(o) + \sum_{j=o}^{t-1} A^j\underline{b}N(t - j), \quad t\epsilon T \tag{3.30}$$

where $A^o = I$ (the identity matrix).

The covariance function of $\{\underline{X}(t), t\epsilon T\}$ is a matrix, and is given by

$$\begin{aligned} K_{\underline{X}}(s, t) &= \text{cov}[\underline{X}(t), \underline{X}(s)] \\ &= \text{cov}[A^t\underline{X}(o) + \sum_{j=o}^{t-1} A^j\underline{b}N(t - j), A^s\underline{X}(o) + \sum_{j=o}^{s-1} A^j\underline{b}N(s - j)] \end{aligned} \tag{3.31}$$

for $s, t\epsilon T$. Since $\{N(t), t\epsilon T\}$ is an independent random sequence, with zero-mean, Equation 3.31 becomes

$$K_{\underline{X}}(s, t) = \operatorname{cov}[A^t \underline{X}(o), A^s \underline{X}(o)]$$

$$+ \operatorname{cov}[\sum_{j=o}^{t-1} A^j \underline{b} N(t - j), \sum_{j=o}^{s-1} A^j \underline{b} N(s - j)$$

$$= E[A^t (\underline{X}(o) - \underline{m}_{\underline{X}}(o))(\underline{X}(o) - \underline{m}_{\underline{X}}(o))' A'^s]$$

$$+ E[\sum_{j=o}^{t-1} \sum_{i=o}^{s-1} A^j \underline{b}\, \underline{b}' A'^i N(t - j)\, N(s - i)] \qquad (3.32)$$

for $s, t \in T$. Letting $V_o = E[(\underline{X}(o) - \underline{m}_{\underline{X}}(o))']$, we have

$$K_{\underline{X}}(s, t) = A^t V_o (A')^s + \sum_{j=o}^{t-1} \sum_{i=o}^{s-1} A^j \underline{b}\, \underline{b}' A'^i E[N(t - j)\, N(s - i)],$$

$$s, t \in T \qquad (3.33)$$

Using Equation 3.13, we have

$$K_{\underline{X}}(s, t) = A^t V_o (A')^s + \sigma^2 \sum_{j=o}^{t-1} \sum_{i=o}^{s-1} A^j \underline{b}\, \underline{b}'\, A'^i \delta(t - j, s - i)$$

$$= A^t V_o (A')^s + \sigma^2 \sum_{j=o}^{\min(t-1, s-1)} A^{|t-s|+j}\, \underline{b}\, \underline{b}' A'^j, \quad s, t \in T$$

$$= A^t V_o (A')^s + \sigma^2 A^{|t-s|} \sum_{j=o}^{\min(t,s)-1} A^j \underline{b}\, \underline{b}' A'^j, \quad s, t \in T$$

(3.34)

Equation 3.34 is the covariance function of the vector sequence $\{\underline{X}(t), t \in T\}$. The $(k, k)^{th}$ entry of this matrix is the covariance function of the random sequence $\{X(t), t \in T\}$.

Note that if A is a convergent matrix, that is, if A has all its eigenvalues within the unit circle (9),

$$K_{\underline{X}}(s, t) \approx \sigma^2 A^{|t-s|} [\sum_{j=o}^{\infty} A^j \underline{b}\, \underline{b}' A'^j], \quad s, t \in T \qquad (3.35)$$

for large s and t. This represents a covariance stationary process.

2. Closed form determination

To determine the covariance function in closed form, we may use the procedure in Appendix B to find h(t), t = 0, 1, 2, · · ·, the discrete filter relating the output y of the filter to its input u, where

$$y(t) = \sum_{j=o}^{t} h(j)\, u(t - j),\ t = 0, 1, 2, \cdots \tag{3.36}$$

If the output of the filter is the random sequence $\{X(t),\ t\epsilon T\}$ and the input is the independent random sequence $\{N(t),\ t\epsilon T\}$ with zero mean, the covariance function of $\{X(t),\ t\epsilon T\}$, using the same procedure in determining Equation 3.34, is given by

$$K_X(s,\ t) = \sum_{j=o}^{\min(s,t)} \sigma^2 h(t - j)\, h(s - j),\ s,\ t\epsilon T \tag{3.37}$$

If the discrete filter is of the form given in Appendix B, Equation B.21, where

$$h(t) = \sum_{i=1}^{k} c_i (\rho_i)^t,\ t = 0, 1, 2, \cdots \tag{3.38}$$

Equation 3.37 becomes

$$K_X(s,\ t) = \sum_{j=o}^{\min(s,\ t)} \sigma^2 \sum_{i=1}^{k} \sum_{m=1}^{k} c_i c_m (\rho_i)^t (\rho_m)^s (\rho_i \rho_m)^{-j}$$

$$= \sigma^2 \sum_{i=1}^{k} \sum_{m=1}^{k} c_i c_m (\rho_i)^t (\rho_m)^s \sum_{j=o}^{\min(s,\ t)} (\rho_i \rho_m)^{-j},\ s,\ t\epsilon T \tag{3.39}$$

If $\rho_i \rho_m \neq 1$,

$$\sum_{j=o}^{\min(s,t)} (\rho_i \rho_m)^{-j} = \sum_{j=o}^{\min(s,t)} \left(\frac{1}{\rho_i \rho_m}\right)^j$$

$$= \frac{\left(\frac{1}{\rho_i \rho_m}\right)^{\min(s,t)+1} - 1}{\left(\frac{1}{\rho_i \rho_m}\right) - 1}$$

$$= \frac{(\rho_i \rho_m)^{-\min(s,t)} - \rho_i \rho_m}{1 - \rho_i \rho_m} \tag{3.40}$$

If $\rho_i \rho_m = 1$,

$$\sum_{j=o}^{\min(s,t)} (\rho_i \rho_m)^{-j} = \min(s,t) + 1 \tag{3.41}$$

Thus, Equation 3.39 becomes

$$K_X(s, t) = \sigma^2 \sum_{i=1}^{k} \sum_{m=1}^{k} c_i c_m (\rho_i)^t (\rho_m)^s \frac{(\rho_i \rho_m)^{-\min(s,t)} - \rho_i \rho_m}{1 - \rho_i \rho_m}, \quad t, s \epsilon T \tag{3.42}$$

where, if a $\rho_i \rho_m = 1$ for some i and m,

$$\lim_{\rho_i \rho_m \to 1} \frac{(\rho_i \rho_m)^{-\min(s,t)} - \rho_i \rho_m}{1 - \rho_i \rho_m} = \min(s, t) + 1 \tag{3.43}$$

Also, if no $\rho_i \rho_m = 1$, and if we let

$$b_{im} = \frac{c_i c_m}{1 - \rho_i \rho_m} = b_{mi}, \quad i, m = 1, \cdots, k \tag{3.44}$$

then

$$K_X(s, t) = \sigma^2 \sum_{i=1}^{k} \sum_{m=1}^{k} b_{im}(\rho_i)^{|t-s|}$$

$$- \sigma^2 \sum_{i=1}^{k} \sum_{m=1}^{k} b_{im} (\rho_i)^{t+1)} (\rho_m)^{s+1}, \; t, \; s \epsilon T \qquad (3.45)$$

If $|\rho_i| < 1$ for all i, Equation 3.45 becomes

$$K_X(s, t) \approx \sigma^2 \sum_{i=1}^{k} \sum_{m=1}^{k} b_{im}(\rho_i)^{|t-s|}, \; t, \; s \epsilon T \qquad (3.46)$$

for large t and s. This represents a covariance stationary process.

An advantage in this closed form method is evident from the fact that the covariance function of Equation 3.34 must be calculated for each s and t in T, resulting in a very large number of arithmetic operations.

IV. NUMERICAL INVESTIGATION

A. Minimization of the Error

As stated in Section B of Chapter III, the problem is to minimize the error $E_{T_1}(v)$ of Equation 3.18 in some sense for all v, with the proper choice of the a_i's and order k.

An average of the square of $E_{T_1}(v)$ was chosen because of simplicity. That is,

$$E = \frac{1}{T_2} \sum_{v \epsilon T_2} [E_{T_1}(v)]^2 \tag{4.1}$$

where T_2 is the range of the summation over v.

E would be minimized if its gradient were zero with $\underline{\nabla}^2_{\underline{\alpha}} E$ positive definite, where $\underline{\nabla}^2_{\underline{\alpha}} E$ is the second directional derivative in the direction $\underline{\alpha}$ normal to $\underline{\nabla} E$ (11). That is, set

$$\underline{\nabla} E = \frac{2}{T_2} \sum_{v \epsilon T_2} E_{T_1}(v) \cdot \underline{\nabla} E_{T_1}(v) = 0 \tag{4.2}$$

Substituting Equation 3.18 into Equation 4.1, we have

$$E = \frac{1}{T_2} \sum_{v \epsilon T_2} \left[\frac{1}{T_1} \sum_{t \epsilon T_1} \sum_{i=o}^{k} \sum_{j=o}^{k} a_i a_j X(t - i)\, X(t - j + v) - K_N(v) \right]^2 \tag{4.3}$$

Equation 4.2 then becomes

$$\frac{\partial E}{\partial a_\alpha} = \frac{2}{T_2} \sum_{v \epsilon T_2} \{[\frac{1}{T_1} \sum_{t \epsilon T_1} \sum_{i=o}^{k} \sum_{j=o}^{k} a_i a_j X(t - i)X(t - j + v)$$

$$- K_N(v)] \cdot [\frac{1}{T_1} \sum_{t \epsilon T_1} \sum_{i=o}^{k} a_i [X(t - i)X(t - \alpha + v)$$

$$+ X(t - \alpha)X(t - i + v)]]\}$$

$$= 0, \alpha = 0, 1, \cdots, k \qquad (4.4)$$

Equation 4.4 is not solvable in closed form, thus a computer gradient technique is necessary. A steepest descent method was chosen, which is described in Appendix A.

Essentially, the steepest descent technique uses Equations 4.3 and 4.4 to minimize the error function for a given k. However, it is not feasible to compute the time average each time the error and its gradient are computed. This problem can be eliminated by interchanging the summations in Equation 4.3. That is, let

$$M_{T_1 ij}(v) = \frac{1}{T_1} \sum_{t \epsilon T_1} X(t - i)X(t - j + v), i,j = 0, 1, \cdots, k; \quad v \epsilon T_2 \qquad (4.5)$$

be a finite-time average, which can be computed separately as a three-dimensional array independent of the a_i's. Equation 4.3 and Equation 4.4 become respectively,

$$E = \frac{1}{T_2} \sum_{v \epsilon T_2} [\sum_{i=o}^{k} \sum_{j=o}^{k} a_i a_j M_{T_1 ij}(v) - K_N(v)]^2 \qquad (4.6)$$

and

$$\frac{\partial E}{\partial a_\alpha} = \frac{2}{T_2} \sum_{v \epsilon T_2} \{[\sum_{i=o}^{k} \sum_{j=o}^{k} a_i a_j M_{T_1 ij}(v) - K_N(v)]$$

$$\cdot [\sum_{i=o}^{k} a_i [M_{T_1 i\alpha}(v) + M_{T_1 \alpha i}(v)]]\}$$

$$= 0, \alpha = 0, 1, \cdots, k \tag{4.7}$$

Due to the finite length of record of the measured sequence $\{X(t), t \epsilon T\}$, a criterion is necessary for the selection of the ranges T_1 and T_2, and how they fit along with the order k, into the record length T_3. The criterion chosen is shown in Figure 4.1.

It is usually desirable for T_2, the range of v, to be no longer than 5 to 10% of T_1 (2). In this case, it was chosen even shorter to save computer time. There was no loss of accuracy in this because the covariance function being estimated was that of a "white noise" sequence, which has a very short correlation time.

As shown in Figure 4.1, in order to compute the "sliding" averages of Equation 4.5, the record length T_3 is related to T_2, T_1, and k by

$$T_3 = T_1 + T_2 + k \tag{4.8}$$

Thus, the summation on t in Equations 4.3, 4.4 and 4.5 will be from $t = k + 1$ to $t = T_1 + k$, and the summation on v in Equations 4.1 through 4.7 will be from $v = T_1 + k + 1$ to $v = T_1 + k + T_2$.

B. Numerical Verification

In order to verify the theory in Chapter III as well as the computer programs, a simulated sequence with known a_i's was generated. A simple random walk was selected. That is, a sequence $\{X(t), t \epsilon T\}$ represented

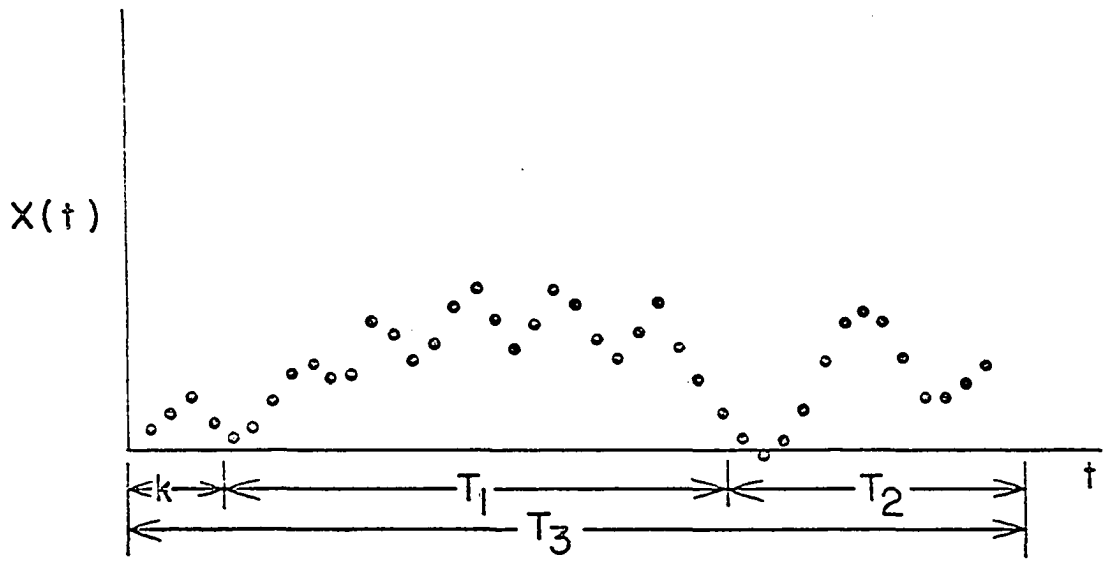

Figure 4.1. Selection of the ranges of T_1, T_2, and k

by the difference equation

$$X(t) = X(t - 1) + N(t), \quad t\epsilon T \tag{4.9}$$

was generated, where $\{N(t), t\epsilon T\}$ was a sequence of independent Gaussian random variables with zero mean and unity variance. This sequence was chosen from a table of Gaussian random variables (25).

In this simulation, T_1 was chosen to be 950, and T_2 was 49. Thus $T_3 = 1000$. The steepest descent program was used to determine how accurately we could converge to the true values of a_o and a_1, which were 1 and - 1, not necessarily respectively.

Using $K_N(v) = \delta(v, o)$, the covariance of input sequence used in Equations 4.6 and 4.7, the steepest descent program converged to coefficients that were about 5% in error. This was understandable since the absolute values of the estimated correlations of the chosen random variables were as high as 0.06.

In looking at Equation 3.18, we can see that $\{N(t), t\epsilon T\}$ need not be an independent sequence, but only a covariance stationary ergodic sequence, as long as $K_N(v)$ is known. If this is so, the procedure of Section B of Chapter III becomes one of system identification.

In this case $K_N(v)$ was not known exactly, but could be estimated as in Equation 3.17. This estimated covariance function was used in Equations 4.6 and 4.7. The steepest descent program then converged to coefficients that were less than 0.2% in error. Thus, the accuracy of the steepest descent program was verified.

All computation was performed on an IBM 360/50 computer system with double precision arithmetic.

V. APPLICATION TO ACTUAL GYRO DRIFT RATE DATA

A. History

The problem of modeling gyro drift rate has probably been one of the more troublesome problems in the field of inertial navigation. Previous papers on this subject have dealt with this problem in different ways.

Hammon (14, 15) assumed that the gyro drift rate was a covariance stationary process, and estimated, by means of finite-time averages, an ensemble of autocorrelation functions, and then calculated the ensemble average of these finite-time averages. His result was a covariance function of the type

$$K_X(v) = A_1^2 e^{-c_1|v|} + A_2^2 e^{-c_2|v|} \tag{5.1}$$

where $c_1 \gg c_2$. The fallacy in this procedure is that time averages are meaningless if the process is nonstationary.

Dushman (8) used the same procedure as Hammon when he assumed a covariance stationary process, but when he assumed a nonstationary process, he used an ensemble average. In both cases he used a least squares fit to an assumed model. This is a legitimate procedure for a nonstationary process, but has two disadvantages. The first disadvantage is the requirement of an ensemble of sample functions. The second is the restriction imposed by the model he assumed.

Thus, we can see why a new method is desired.

B. Applying the New Model

The procedures discussed in Chapters III and IV were applied to three sets of gyro drift rate data. One set of data represents a gyro of a certain model number, and the other two sets of data represent a

different gyro of the same model number on two different test runs. These data will be denoted as gyro 1, gyro 2A, and gyro 2B respectively.

These data were furnished by the A-C Electronics Division of the General Motors Corporation. For security reasons, no amplitude scaling was furnished, therefore the amplitudes in the results are only relative amplitudes. The procedure in obtaining this data as well as the model number of the gyros was not made known to the author.

The procedure used to analyze each set of data was the same, except for a minor difference in the length of the averaging time T_1. Table 5.1 illustrates the parameters described in Chapters III and IV.

Table 5.1. Parameters used in computation of models

Gyro	Maximum order k	Averaging time T_1	Correlation range T_2	Scale factor of time d_t	Variance of N(t) σ^2
1	6	3600	150	103 sec	1
2A	6	3800	150	103 sec	1
2B	6	3700	150	103 sec	1

The results of this procedure are given in Tables 5.2 through 5.6 and Figures 5.1 through 5.18. Table 5.2 illustrates the coefficients $(a_o, \cdots, a_k)$ of Equation 3.5 for each order k, and the estimated variance of N(t), $\hat{K}_{T_1 N}(o)$, of Equation 3.18 for each order k.

Table 5.3 illustrates the coefficients $(a_{k1} \cdots a_{kk}, b_k)$ of

Table 5.2. Coefficients of the difference equations and variance of $\{N(t), t\epsilon T\}$

GYRO	ORDER k	COEFFICIENTS ($\times 10^{-3}$) of Equation 3.5 a_0	a_1	a_2	a_3	a_4	a_5	a_6	ESTIMATED VARIANCE $\hat{R}_{T_1N}(o)$
GYRO 1	1	-6.030337	6.052269						0.7687683
	2	-7.378874	4.400528	2.961585					0.8473576
	3	-8.041593	3.411092	2.734119	1.918171				0.8909514
	4	-8.377259	2.659770	2.508298	1.974056	1.256768			0.9067780
	5	-8.564317	2.472354	2.327785	1.835641	1.184016	0.733795		0.9216001
	6	-8.619262	1.695929	2.480060	2.005418	1.347830	0.894615	0.216904	0.9171443
GYRO 2A	1	-4.945278	4.950623						0.7813316
	2	-5.870314	3.617453	2.253356					0.8524017
	3	-6.099647	3.073251	2.076643	0.950359				0.8533329
	4	-6.466069	2.530700	2.032015	1.275629	0.628597			0.9154538
	5	-6.504750	2.500743	2.044353	1.258221	0.577624	0.124707		0.9243927
	6	-6.443399	2.420226	1.959986	1.277867	0.656700	0.112537	0.016977	0.9038569
GYRO 2B	1	-4.119961	4.131025						0.8061708
	2	-4.787996	3.070820	1.707786					0.8788611
	3	-5.055638	2.747711	1.576405	0.723162				0.9172960
	4	-5.258580	2.274809	1.450335	0.951224	0.575211			0.9479334
	5	-5.283877	2.203978	1.284865	0.904790	0.581298	0.302893		0.9457614
	6	-5.27282	2.181925	1.234312	0.876286	0.554348	0.270784	0.149469	0.9377204

Table 5.3. Entries of transition matrix and input vector, and average error

GYRO	ORDER k	ENTRIES OF A MATRIX AND $\underline{b}$ VECTOR a_{k1}	a_{k2}	a_{k3}	a_{k4}	a_{k5}	a_{k6}	$b_k(\times 10^3)$	ERROR $E(\times 10^3)$
GYRO 1	1	1.003637						0.145397	1.654418
	2	0.596368	0.401360					0.124751	1.018298
	3	0.424181	0.339997	0.238531				0.117378	0.732370
	4	0.317499	0.299418	0.235645	0.150020			0.113671	0.611820
	5	0.288681	0.271801	0.214336	0.138250	0.085681		0.112093	0.543228
	6	0.196762	0.287734	0.232667	0.156374	0.103793	0.025167	0.111109	0.539170
GYRO 2A	1	1.001081						0.178742	1.557472
	2	0.616228	0.383890					0.157275	0.895983
	3	0.503840	0.340453	0.155806				0.151445	0.664902
	4	0.391381	0.314258	0.197280	0.097215			0.147971	0.538518
	5	0.384449	0.314286	0.193431	0.088800	0.017172		0.147808	0.538091
	6	0.375613	0.304185	0.198322	0.101918	0.017465	0.002635	0.147549	0.537201
GYRO 2B	1	1.002685						0.217932	1.330861
	2	0.641358	0.356681					0.195797	0.786405
	3	0.543495	0.311817	0.143042				0.189443	0.568196
	4	0.432589	0.275800	0.180888	0.109385			0.185149	0.449188
	5	0.417113	0.243167	0.171236	0.110014	0.057324		0.184051	0.425015
	6	0.413806	0.234090	0.166189	0.105133	0.0513546	0.028347	0.183651	0.414403

Equations 3.8 and 3.9 for each order k, where the variance of the sequence $\{N(t), t\epsilon T\}$ is unity. Also shown is the average error as given in Equation 4.1.

Table 5.4. Discrete filters for gyro 1

ORDER k	RESPONSE OF H TO $\delta(i, o)$ $h(t) = \sum_{i=o}^{\infty} h(t - i)\,\delta(i, o)$ $(x10^5)$ $t = 0, 1, 2, 3, \cdots$
1	-14.54 exp(.00363t)
2	-8.894 exp(-.00162t)-3.581 exp(-.9113t)cosπt
3	-6.471 exp(.00149t)+5.267 exp(-.7174t)cos(2.204t+3.133)
4	-5.129 exp(.00116t)-2.122 exp(-.6791t)cosπt +4.119 exp(-.6095t)cos(1.734t-3.103)
5	-4.555 exp(-.00051t)+3.508 exp(-.6398t)cos(2.556t+3.126) +3.148 exp(-.5886t)cos(1.418t-3.109)
6	-4.030 exp(.00090t)-.709 exp(-1.0724t)cosπt +3.794 exp(-.7204t)cos(2.406t-2.855) +2.787 exp(-.5860t)cos(1.335t-2.944)

In Tables 5.4 through 5.6 we have the discrete filters representing each gyro for each order k. The coefficients $(a_o, \cdots, a_k)$ of Equation 3.5 used in Equation B.14 of Appendix B are normalized by dividing each computed a_i by the estimated standard deviation $[\hat{K}_{T_1N}(o)]^{\frac{1}{2}}$.

Figures 5.1 through 5.18 are the estimated covariance functions $\hat{K}_{T_1N}(\tau)$, $\tau\epsilon T_2$, for each gyro and each order k.

Table 5.5. Discrete filters for gyro 2A

ORDER k	RESPONSE OF H TO $\delta(i, o)$ $h(t) = \sum_{i=o}^{\infty} h(t - i)\,\delta(i, o)$ $(x10^5)$ $t = 0, 1, 2, \cdots$
1	-17.88 exp(.00108t)
2	-11.365 exp(.000085t)-4.362 exp(-.9575t)cosπt
3	-9.168 exp(.000060t)+5.980 exp(-.9296t)cos(2.251t-3.112)
4	-7.398 exp(.000068t)-2.582exp(-.8190t)cosπt +4.820 exp(-.7560t)cos(1.750t-3.113)
5	-7.231 exp(.000067t)+3.644 exp(-1.115t)cos(2.688t-2.946) +4.057 exp(-0.8618t)cos(1.602t-2.941)
6	-7.062 exp(.000066t)-2.516 exp(-.9572t)cosπt +5.057 exp(-.8611t)cos(1.642t-2.931) +4.260 exp(-1.630t)cos(2.048t+2.145)

Table 5.6. Discrete filters for gyro 2B

ORDER k	RESPONSE OF H TO $\delta(i, o)$ $h(t) = \sum_{i=o}^{t} h(t - i)\,\delta(i, o)$ $(x10^5)$ $t = 0, 1, 2, 3, \cdots$
1	-21.8 exp(.002685t)
2	-14.421 exp(-.00145t)-5.159exp(-1.0295t)cosπ
3	-11.844 exp(-.00103t)+7.105 exp(-.9718t)cos(2.217t+3.104)
4	-9.409 exp(-.00068t)-3.034 exp(-.7853t)cosπt +6.108 exp(-.7135t)cos(1.684t+3.032)
5	-8.572 exp(-.00053t)+4.752 exp(-.7735t)cos(2.556t+3.073) +5.145 exp(-.6557t)cos(1.390t+2.999)
6	-8.231 exp(-.000483t)-2.078 exp(-.7416t)cosπt +4.116 exp(-.7418t)cos(2.116t+2.940) +4.058 exp(-.6688t)cos(1.186t+3.012)

C. Discussion of Results

The goodness of each model for a particular gyro is best shown on Figures 5.1 through 5.18. That is, if the estimated covariance function $\hat{K}_{T_1 N}(\tau)$, $\tau \epsilon T_2$, is satisfactory for the applicable engineering problem, the model which produced this approximately independent random sequence is satisfactory.

We can see from the figures that if the covariance at $\tau = 1$ is not critical, the first order model for all three gyros is quite satisfactory. However, if it is critical, we may need a model of as high as fifth order to bring the covariance at $\tau = 1$ down to an acceptable value.

For the sample sizes T_1 of Table 5.1, the probabilities, P_ϵ that the absolute value of the correlations exceed ϵ, as given in Equation 3.21, for $\epsilon = .1$ and $\epsilon = .05$ are respectively, nearly zero (less than .0001) and approximately .0025. Thus, with this in mind, we can see we should have all applicable correlations in the range $-.1 \leq \rho(\tau) \leq .1$, and a high percentage of the applicable correlations in the range $-.05 \leq \rho(\tau) \leq .05$. In fact, we should not have any correlations very far out of the range $-.05 \leq \rho(\tau) \leq .05$. It can be seen in Figures 5.1 through 5.18 that this is so for all τ for the higher order models, and this is so for $\tau > 1$ in even the first order models.

In Tables 5.4 through 5.6, we see that the even order models have a frequency component of π radians per unit of time. This results from a negative eigenvalue of the A matrix.

In all cases, we have an exponential term which has a relatively small exponent, either positive or negative. This has a strong resemblance to

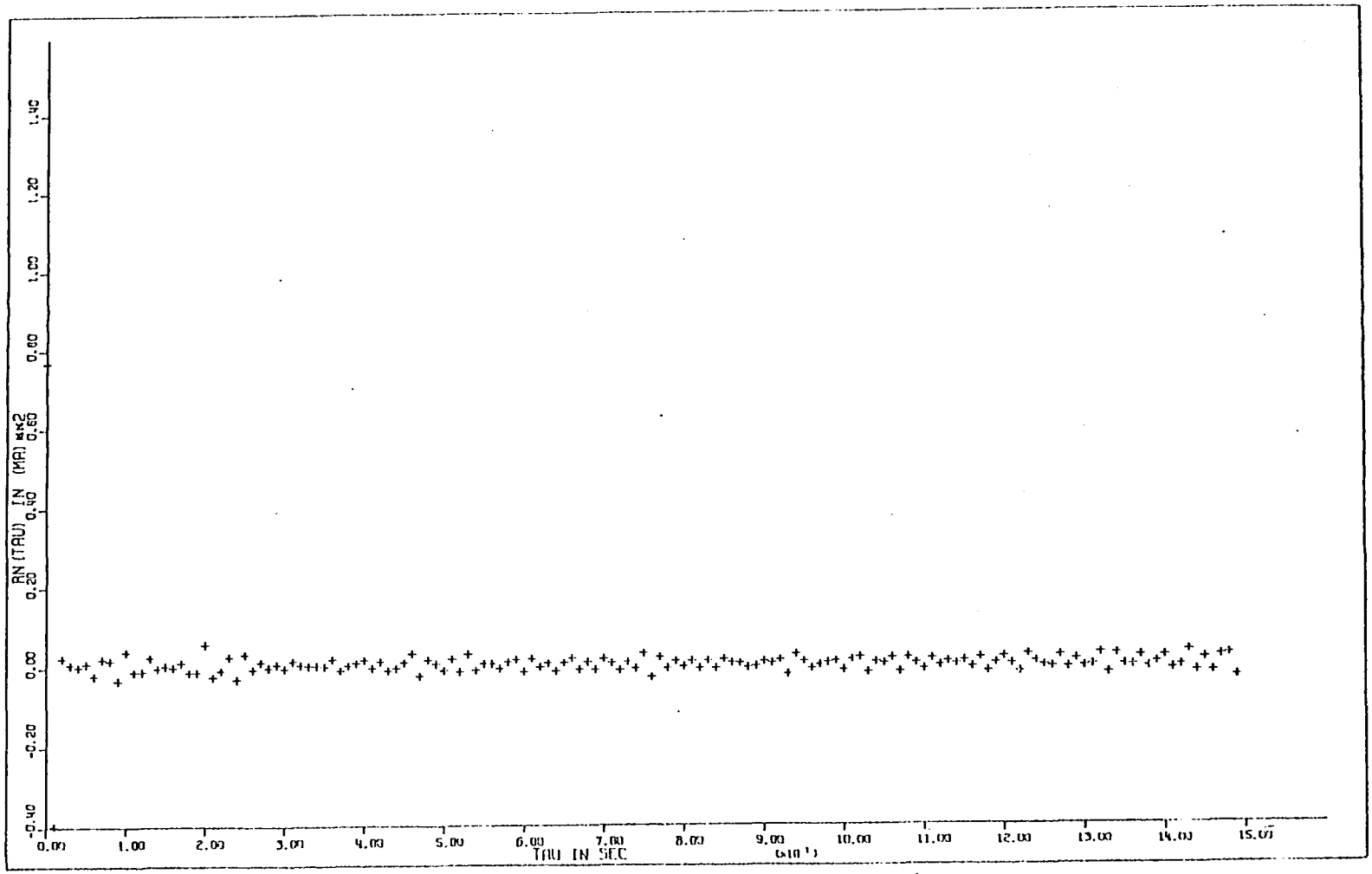

Figure 5.1. Estimated covariance function of {N(t), teT} for 1st order model of gyro 1

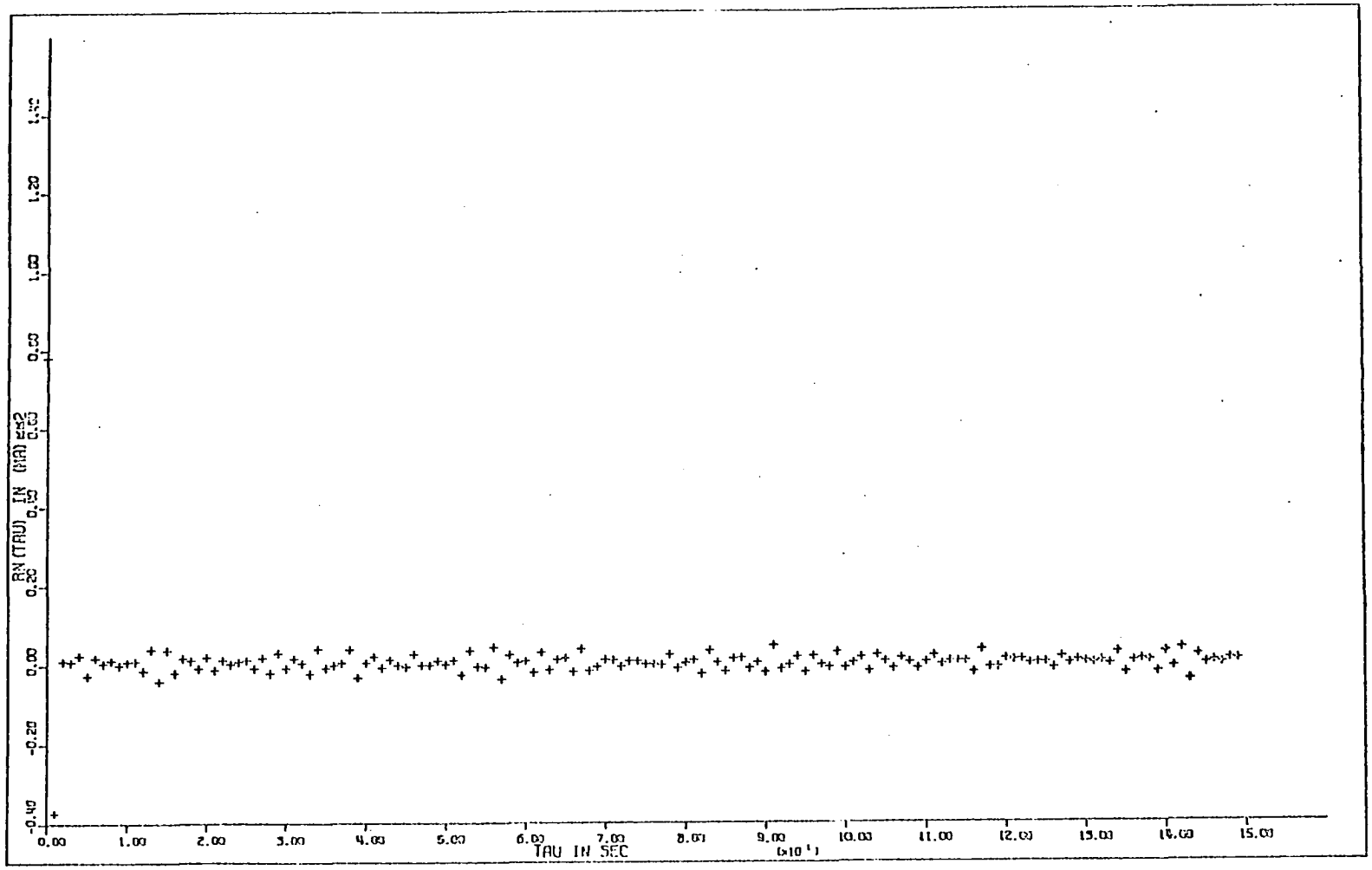

Figure 5.2. Estimated covariance function of $\{N(t),\ t\epsilon T\}$ for 1^{st} order model of gyro 2A

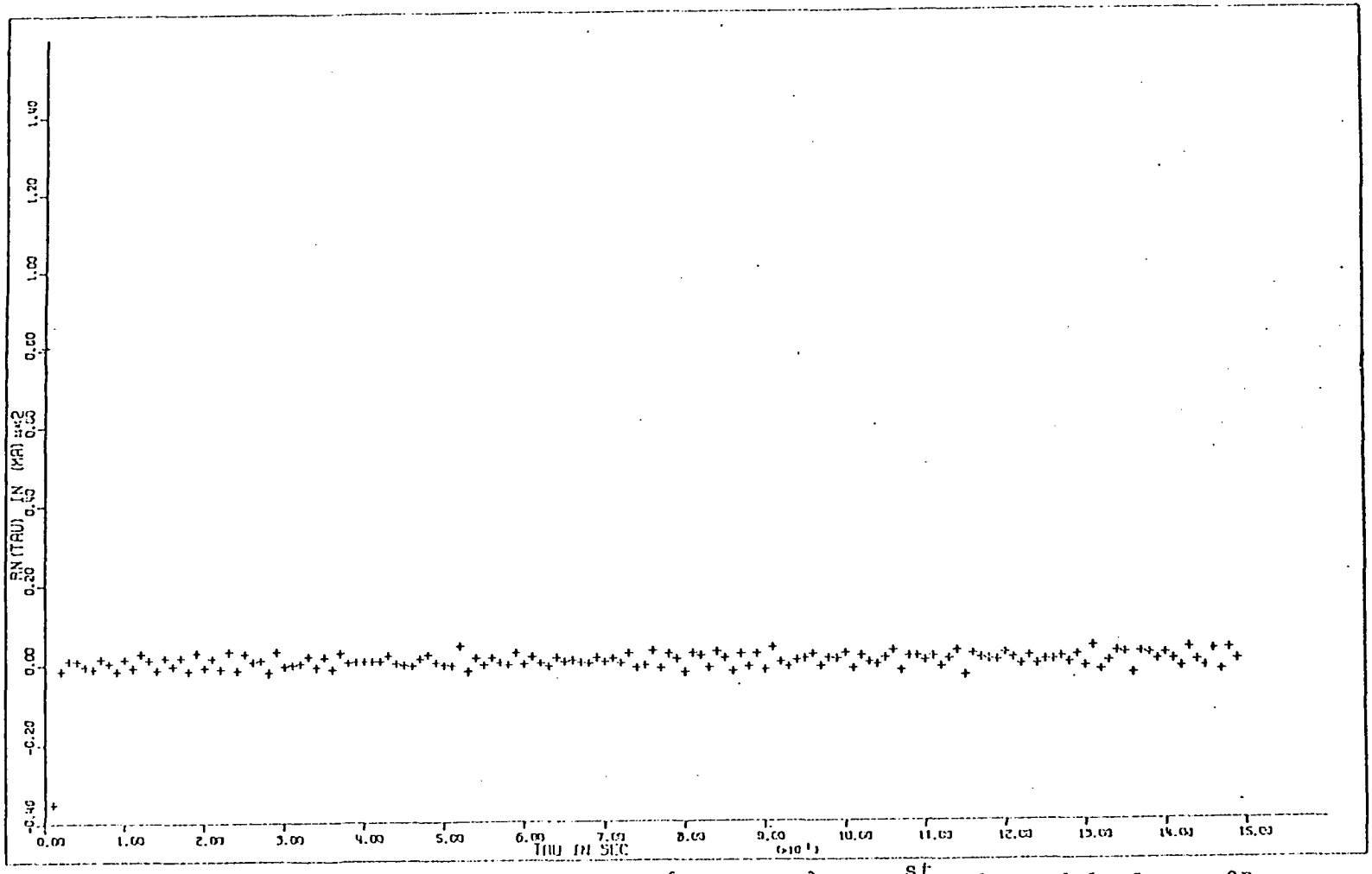

Figure 5.3. Estimated covariance function of $\{N(t), t\epsilon T\}$ for 1^{st} order model of gyro 2B

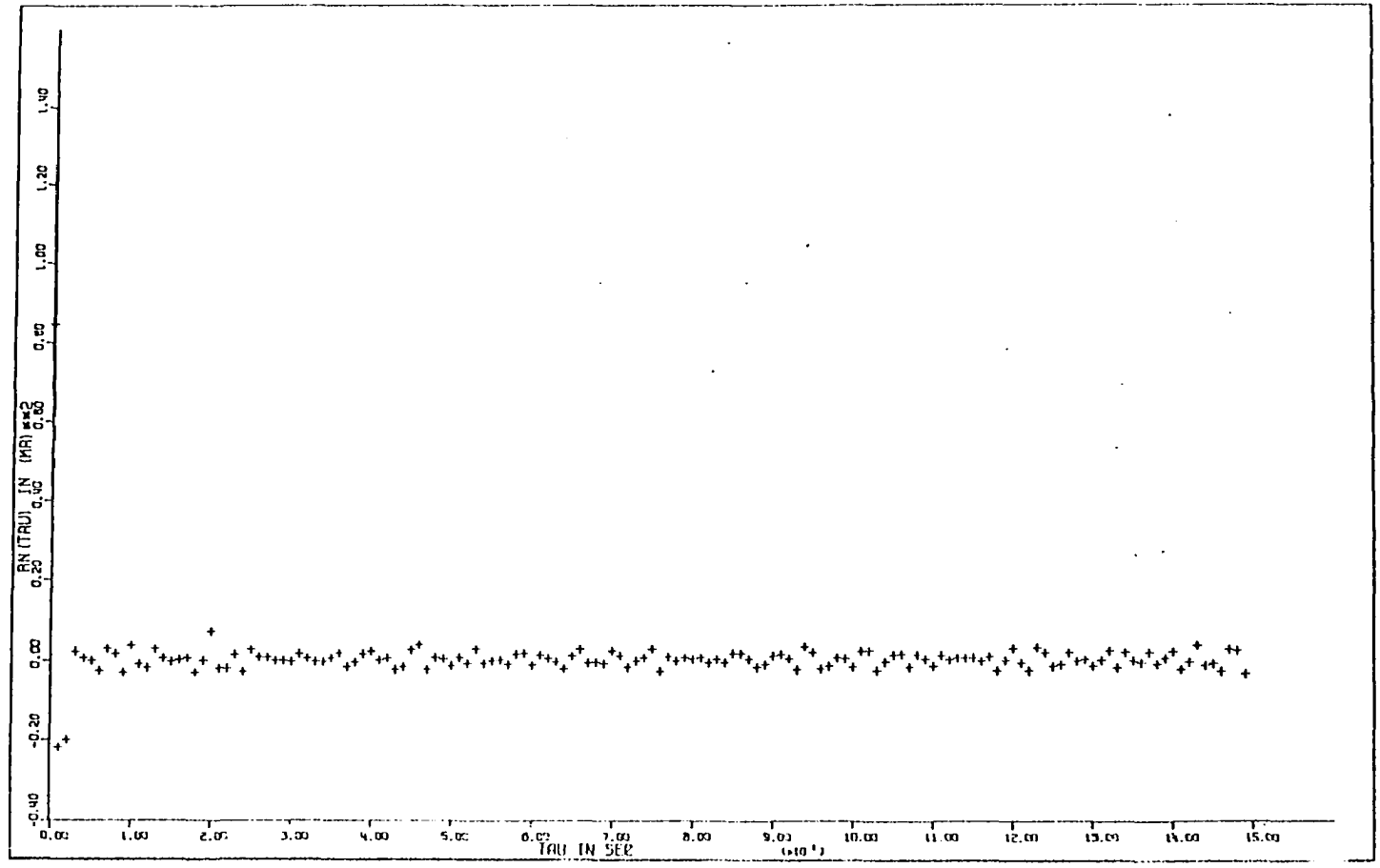

Figure 5.4. Estimated covariance function of $\{N(t),\ t\epsilon T\}$ for 2^{nd} order model of gyro 1

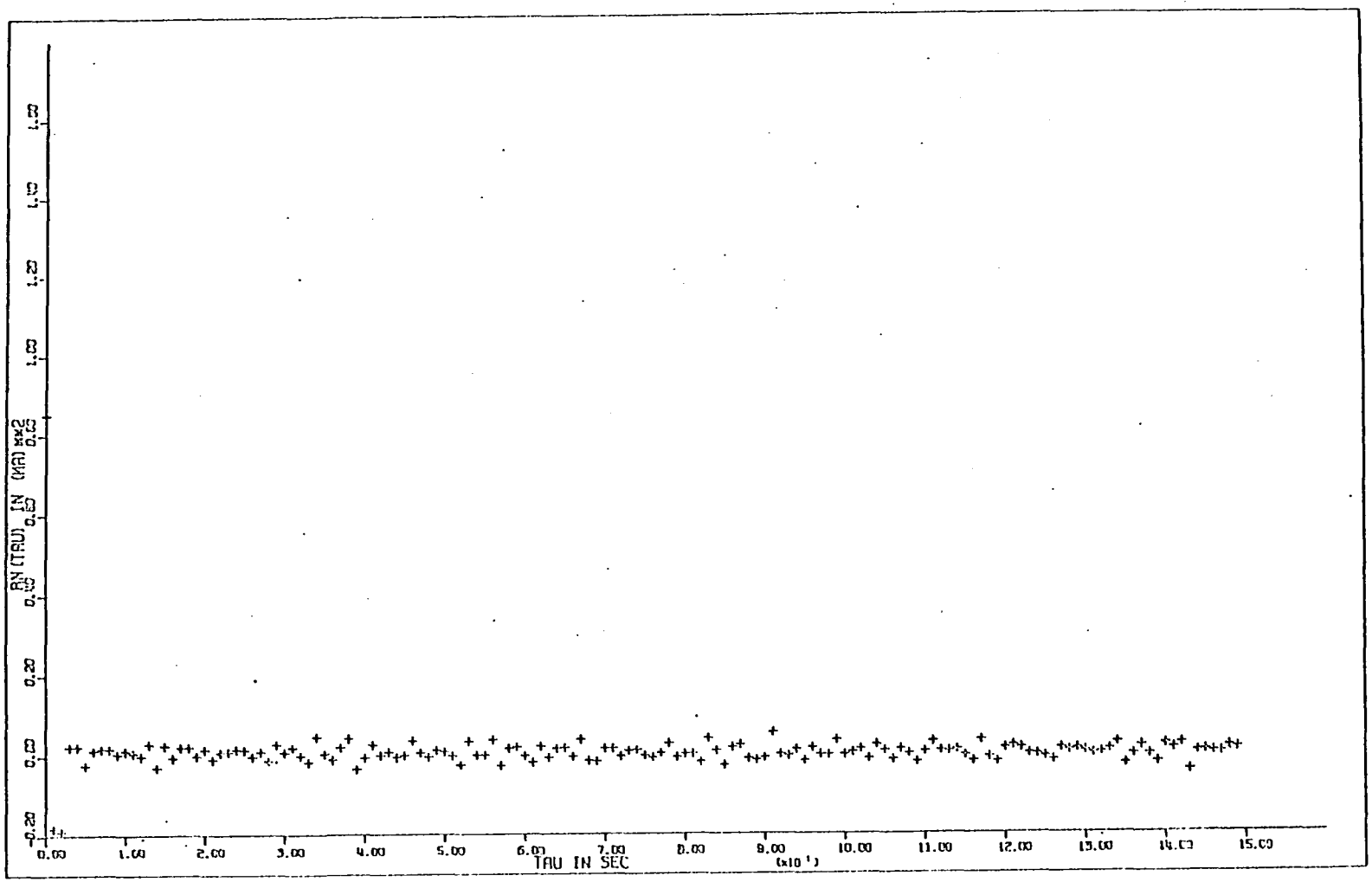

Figure 5.5. Estimated covariance function of {N(t), t∈T} for 2nd order model of gyro 2A

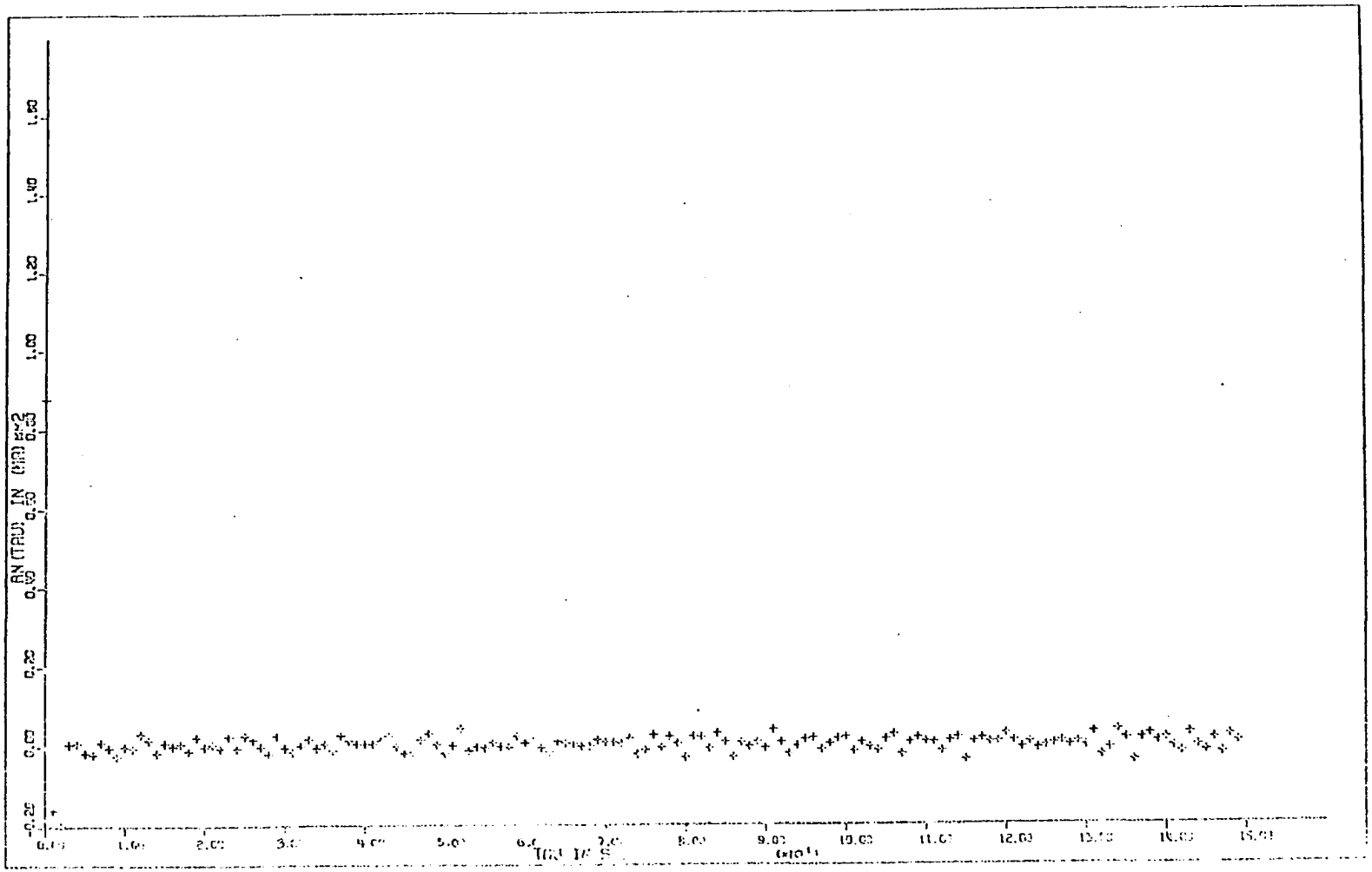

Figure 5.6. Estimated covariance function of {N(t), teT} for 2nd order model of gyro 2B

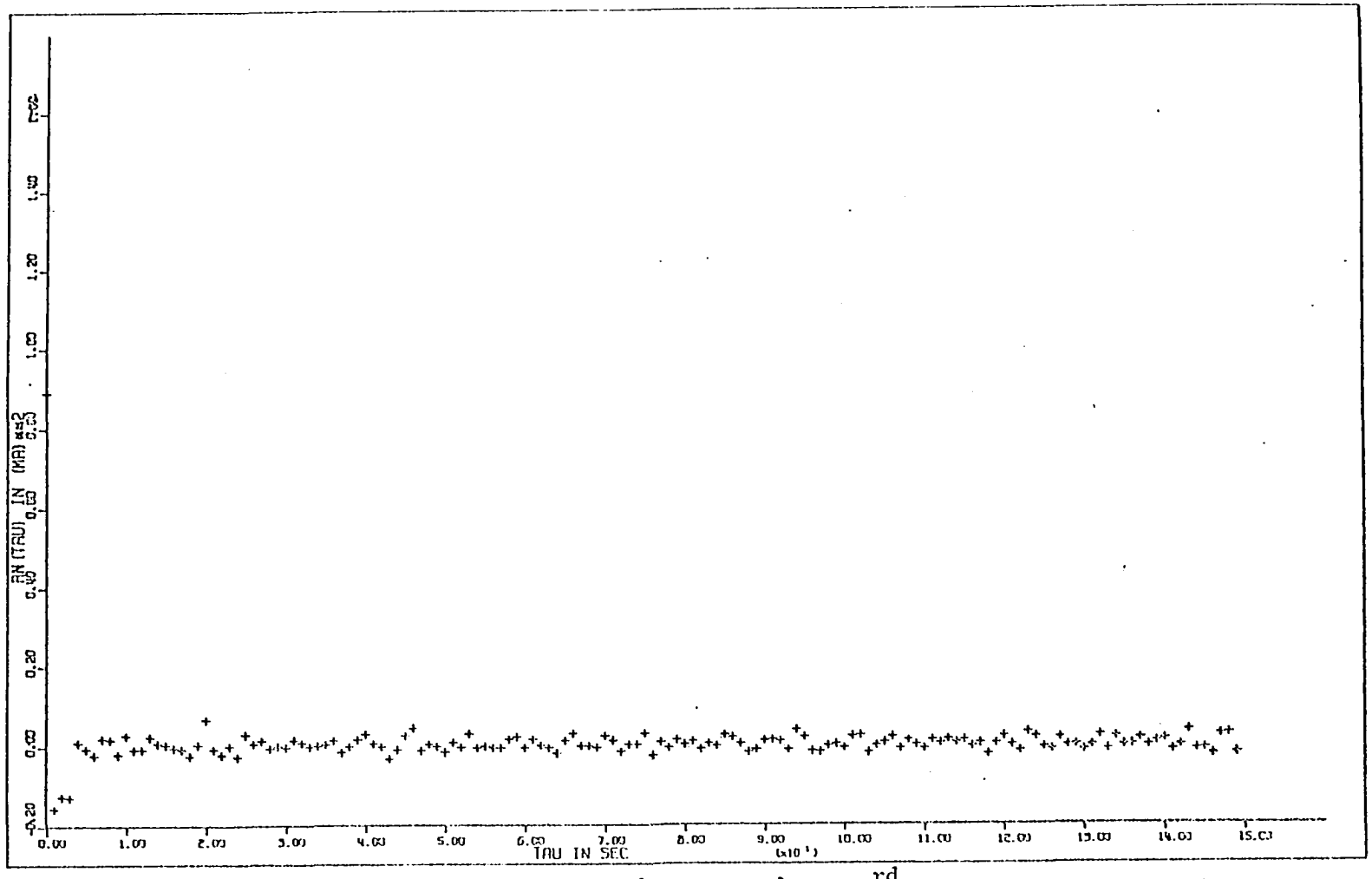

Figure 5.7. Estimated covariance function of $\{N(t),\ t\in T\}$ for 3rd order model of gyro 1

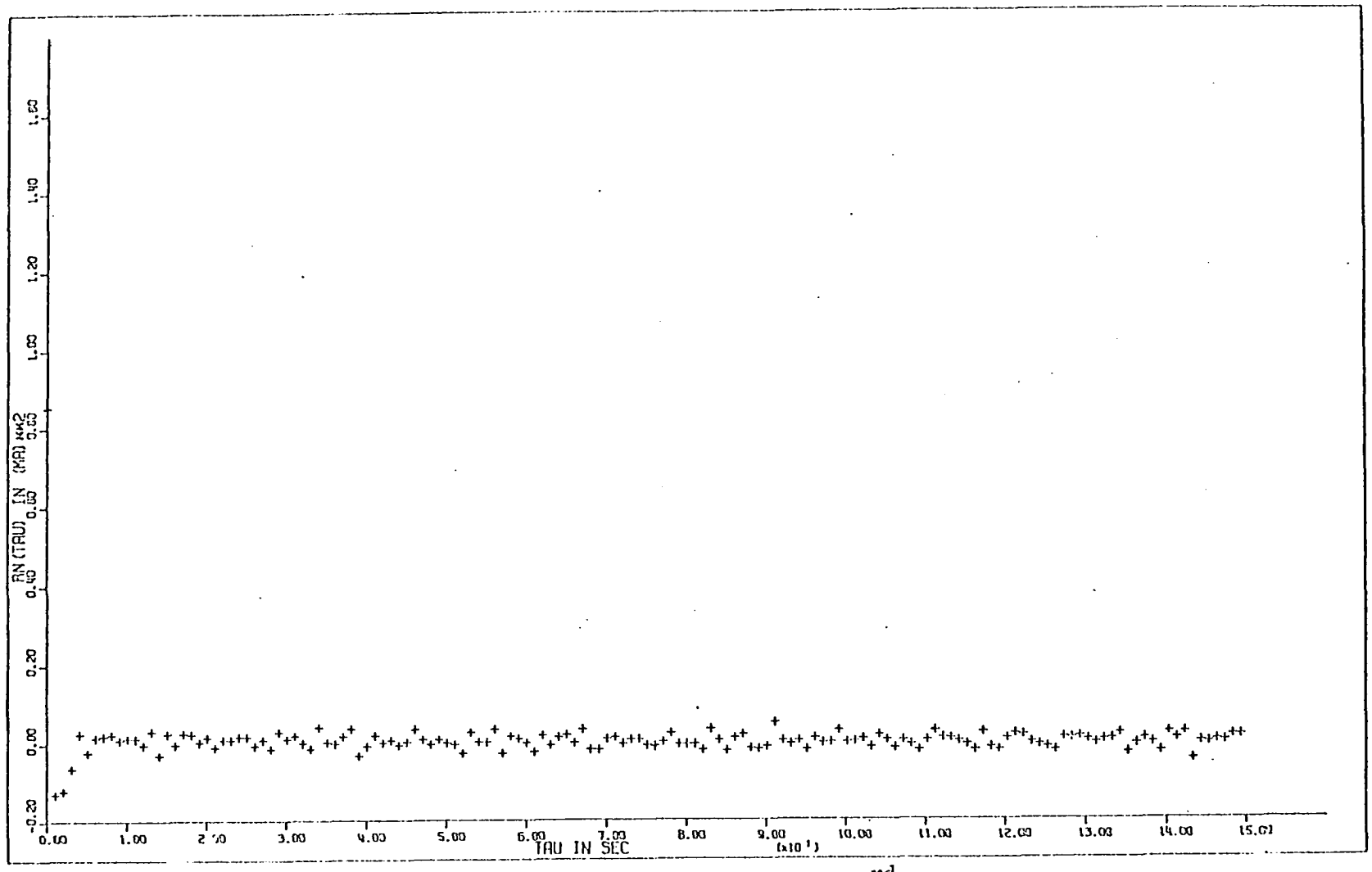

Figure 5.8. Estimated covariance function of {N(t), teT} for 3rd order model of gyro 2A

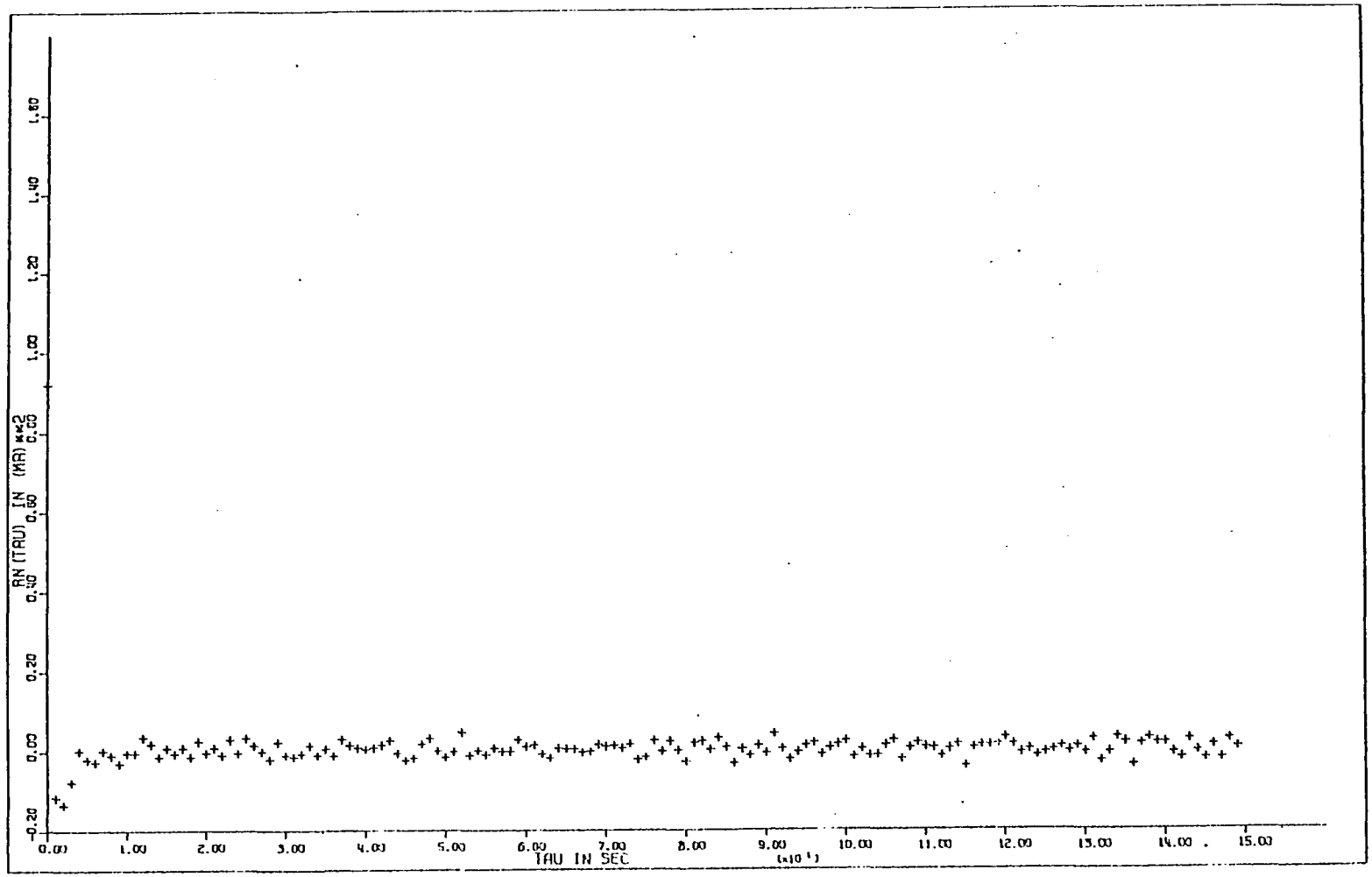

Figure 5.9. Estimated covariance function of $\{N(t),\ t\epsilon T\}$ for 3rd order model of gyro 2B

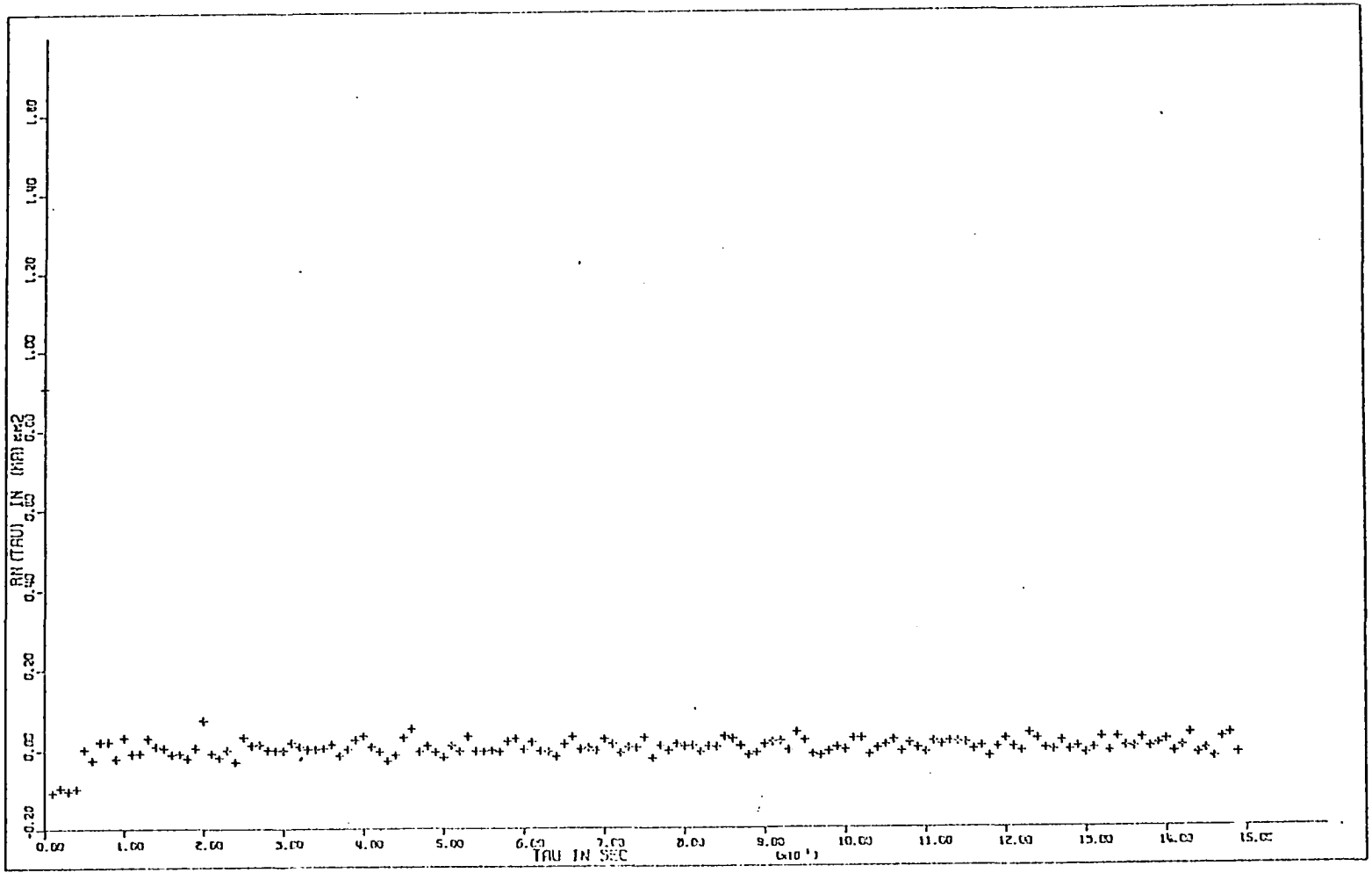

Figure 5.10. Estimated covariance function of $\{N(t),\ t\epsilon T\}$ for 4^{th} order model of gyro 1

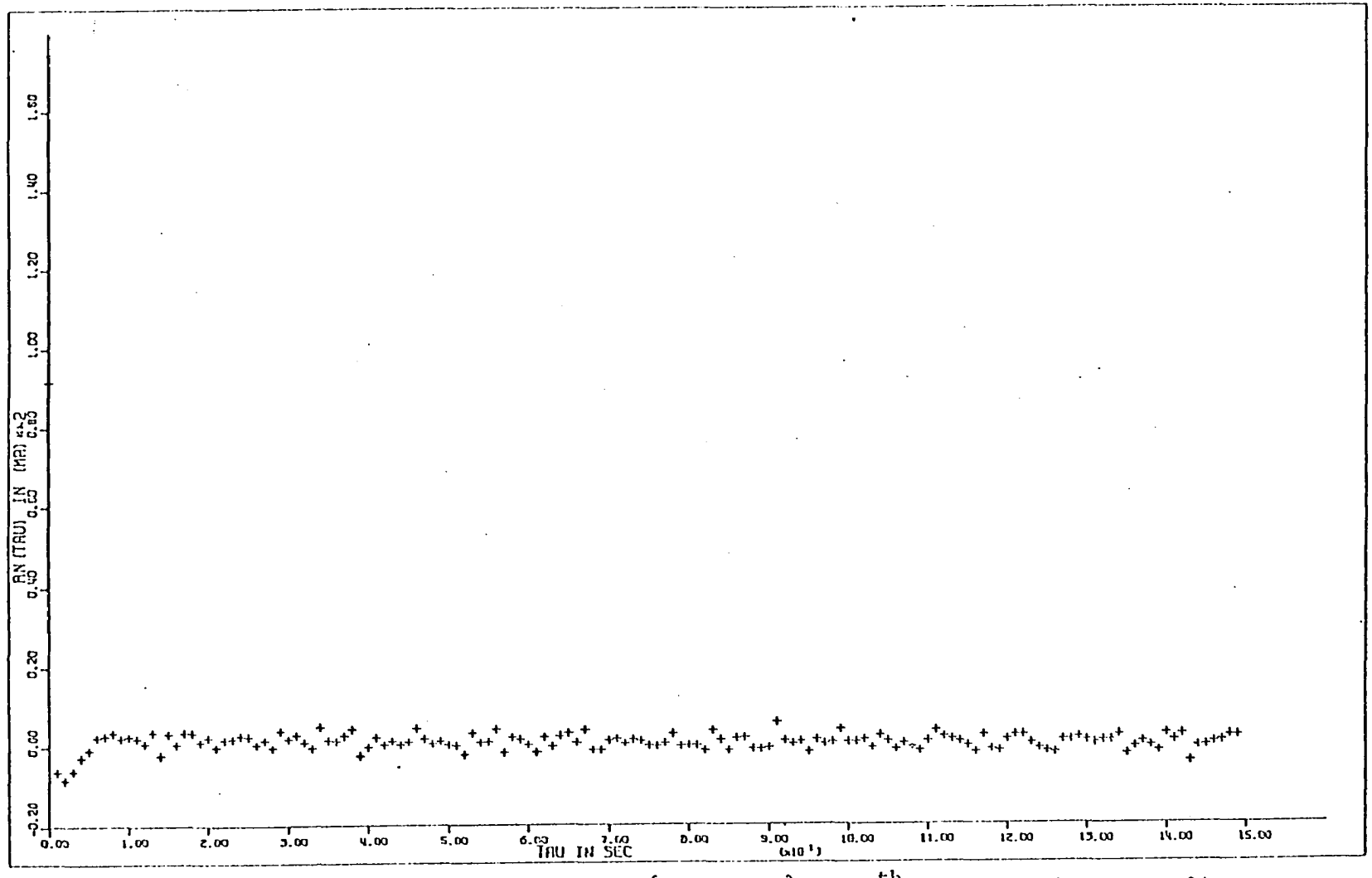

Figure 5.11. Estimated covariance function of $\{N(t),\ t\in T\}$ for 4^{th} order model of gyro 2A

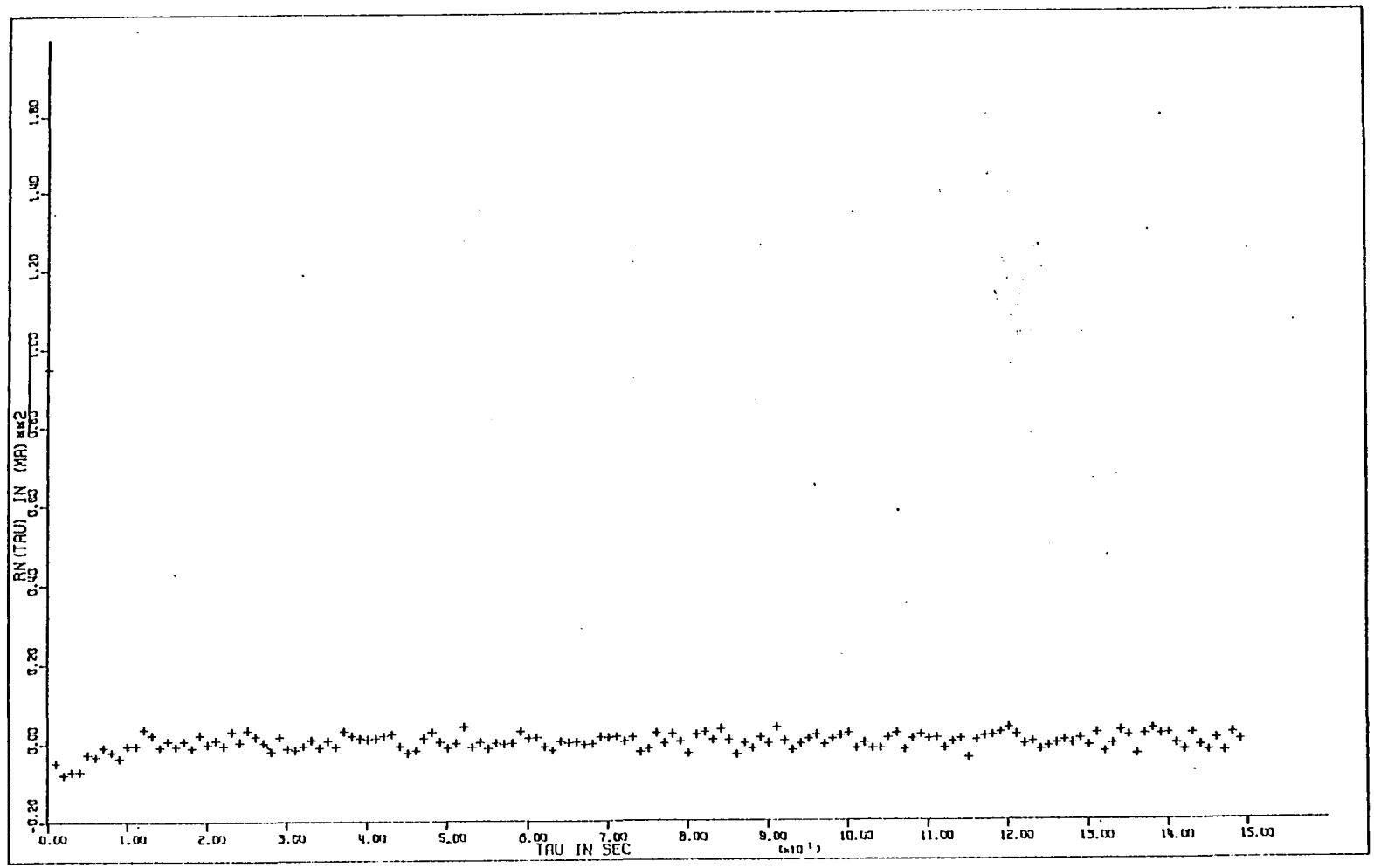

Figure 5.12. Estimated covariance function of {N(t), teT} for 4^{th} order model of gyro 2B

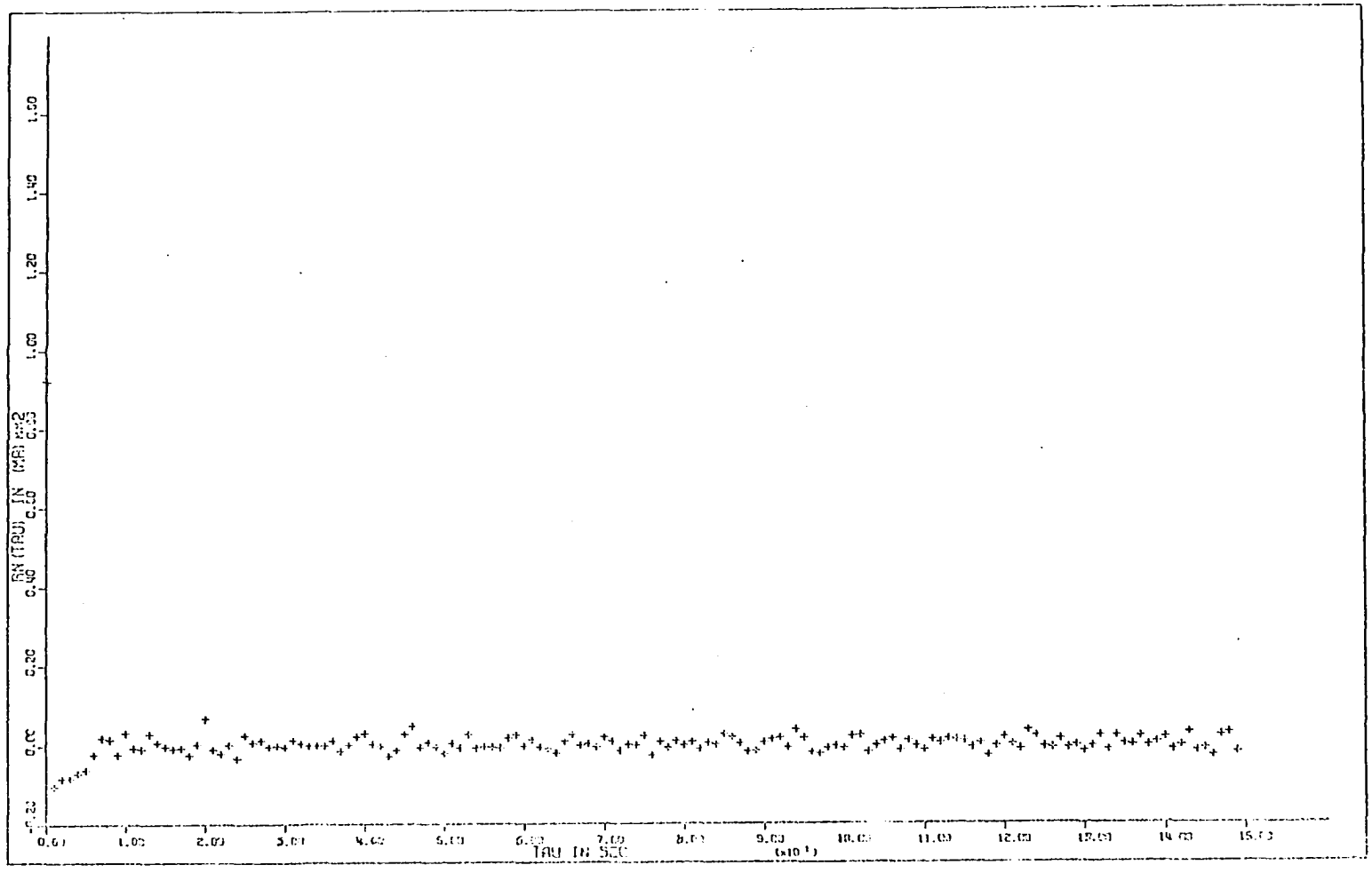

Figure 5.13. Estimated covariance function of {N(t), t∈T} for 5th order model of gyro 1

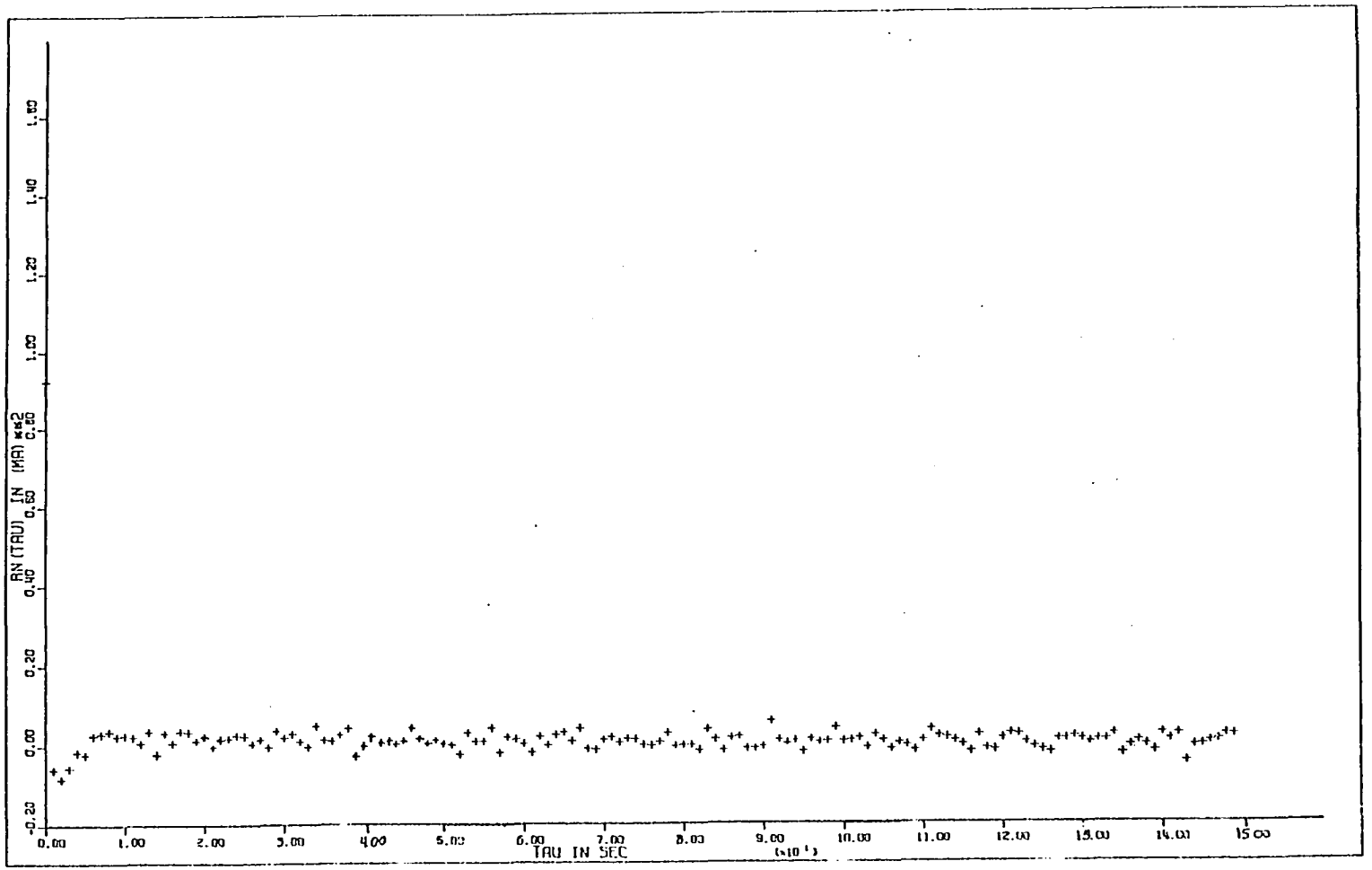

Figure 5.14. Estimated covariance function of $\{N(t),\ t\epsilon T\}$ for 5^{th} order model of gyro 2A

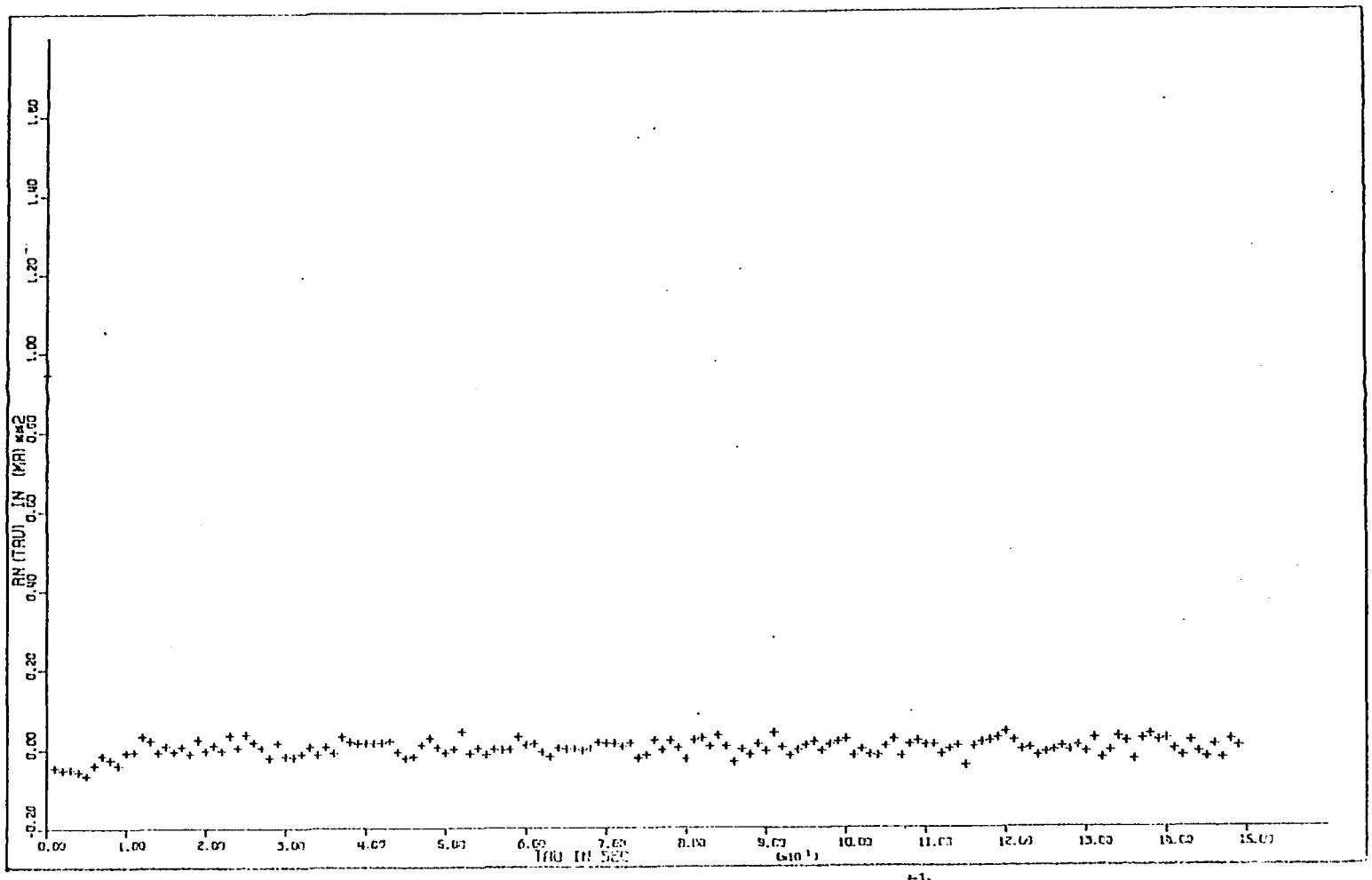

Figure 5.15. Estimated covariance function of $\{N(t),\ t\in T\}$ for 5^{th} order model of gyro 2B

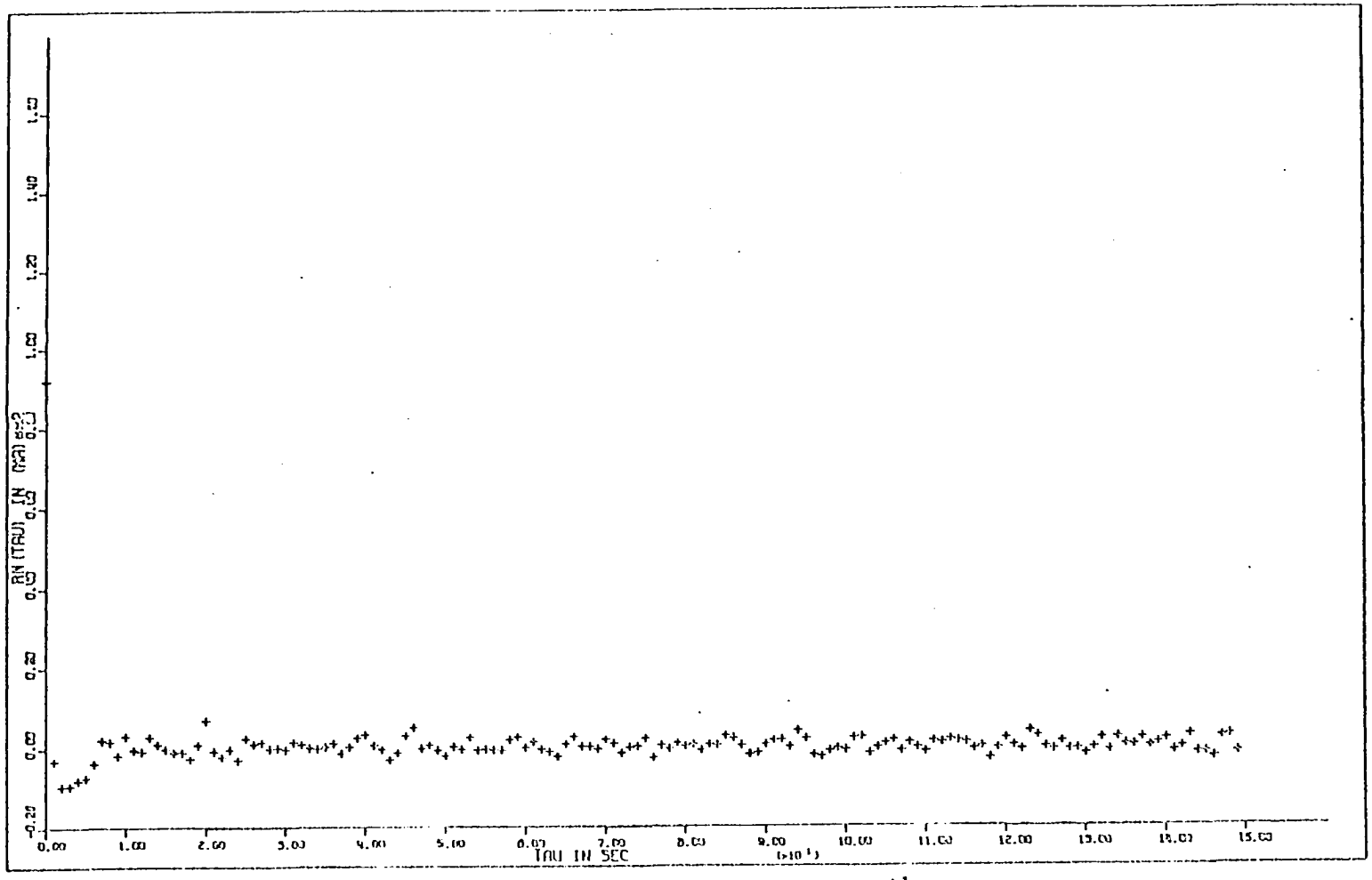

Figure 5.16. Estimated covariance function of {N(t), t∈T} for 6th order model of gyro 1

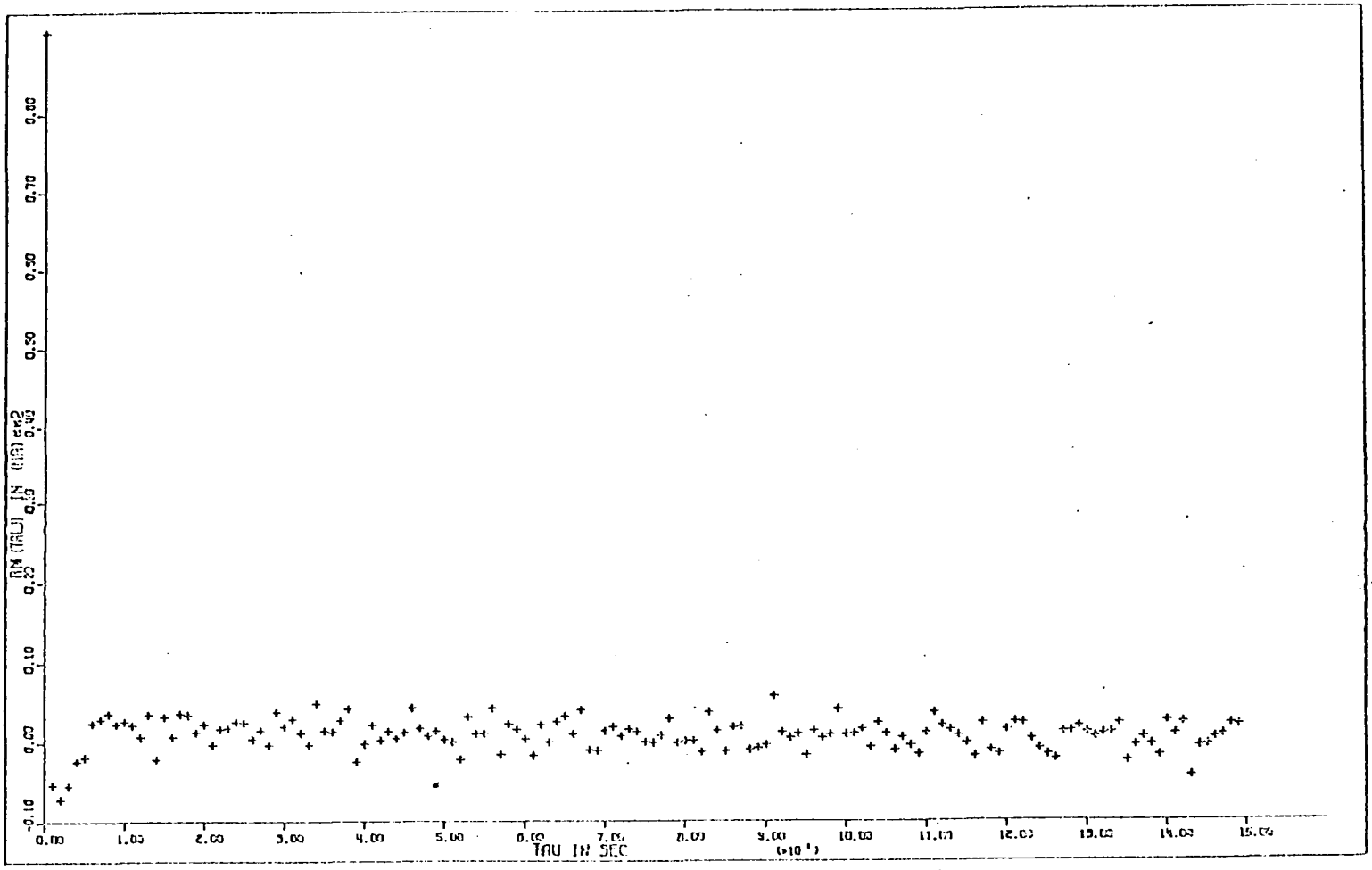

Figure 5.17. Estimated covariance function of $\{N(t),\ t\epsilon T\}$ for 6^{th} order model of gyro 2A

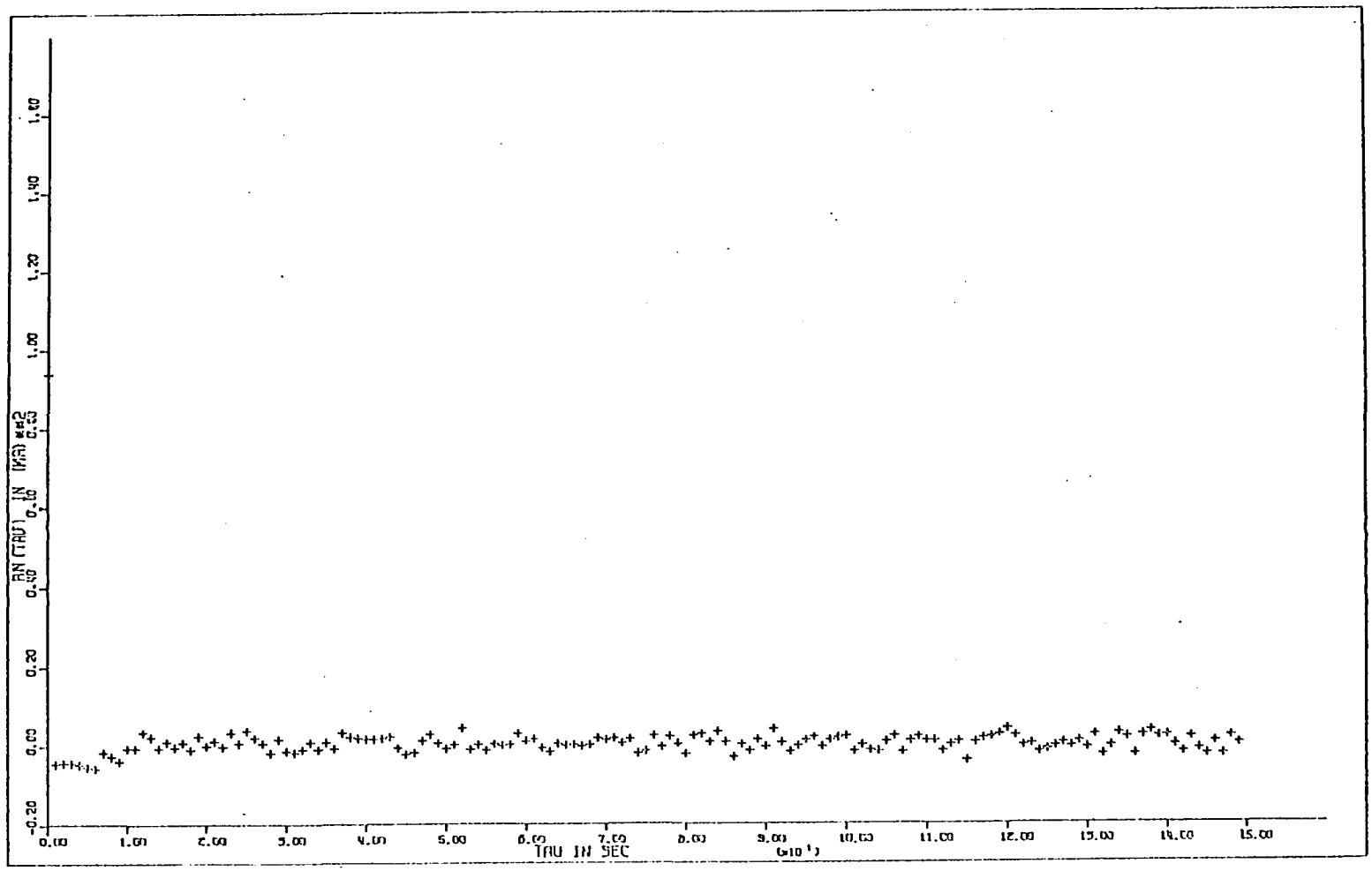

Figure 5.18. Estimated covariance function of {N(t), teT} for 6^{th} order model of gyro 2B

a random walk, which would have a zero exponent.

The exponentials associated with the oscillatory terms have larger negative exponents, indicating correlation between adjoining times in the processes, but very little correlation otherwise.

Thus, we may arrive at the conclusion that all three gyros have a drift rate that is very nearly a random walk with additional correlation between adjoining times, which may be either positive or negative, depending on the model. In many inertial applications, it would appear that a random walk model would be an acceptable model.

VI. SUMMARY AND CONCLUSIONS

A new technique for modeling nonstationary random processes has been introduced and investigated. Under the assumption that the nonstationary process being investigated can be thought of as the result of passing stationary ergodic "white noise" through a time-invariant linear filter, this technique requires no ensemble averaging. Instead, time averages of products of the process at different times are performed. These averages are then used in a steepest descent method to determine the coefficients of a difference equation with a "white noise" sequence forcing function to describe the process at discrete times.

This technique would not be feasible if time averages had to be computed for each iteration of the steepest descent calculations. However, it has been shown that these time averages need be calculated only once, and then stored for use in the steepest descent calculations.

The main advantage in applying this technique rather than that of ensemble averaging is that a large number of sample functions of the process is not required. We may wish to observe the results of more than one sample function, but not the large number required for ensemble averaging.

Another advantage is that only one sample function must be stored in the computer while applying the time averaging technique.

The transformation of the model to the form necessary for augmentation of a system for "Kalman" filtering is also presented. As the difference equation model is only defined for integer values of normalized

time, a technique for extending this model to an interpolated model, from which other discrete models with different time bases can be derived, has been presented in Appendix C. This technique is not unique, and the use of it depends on the applicable engineering problem.

Also presented is a technique for calculating the covariance function of the process being investigated. This covariance function can be calculated from the difference equation model directly, or from a discrete filter describing this difference equation. This last method results in a covariance function in closed form.

The above techniques were applied to three records of gyro drift rate with very satisfactory results. These results are given and discussed in Chapter V. From this analysis, it appears that gyro drift rate can be modeled using the techniques presented.

VII. LITERATURE CITED

1. Alphors, Lars V. Complex analysis. New York, N.Y., McGraw-Hill Book Co., Inc. 1953.

2. Blackman, R. B. and Tukey, J. W. Measurement of power spectra from the point of view of communications engineering. Part I. Bell System Technical Journal 37: 185-282. 1958.

3. Boonton, R. C., Jr. An optimization theory for time-varying linear systems with nonstationary statistical inputs. Institute of Radio Engineers Proceedings 40: 977-981. 1952.

4. Brand, Louis. Differential and difference equations. New York, N.Y., John Wiley and Sons, Inc. 1966.

5. Brown, Robert Grover and Nilsson, James William. Introduction to linear systems analysis. New York, N.Y., John Wiley and Sons, Inc. 1962.

6. Darlington, Sidney. Nonstationary smoothing and prediction using network theory concepts. Institute of Radio Engineers Transactions on Circuit Theory Special Supplement CT-6: 2-13. May 1959.

7. Davis, R. C. On the theory of prediction of nonstationary stochastic processes. Journal of Applied Physics 23: 1047-1053. 1952.

8. Dushman, Allan. On gyro-drift models and their evaluation. Institute of Radio Engineers Transactions on Aerospace and Navigational Electronics ANE-9: 230-234. 1962.

9. Finkbeiner, Daniel T., II. Introduction to matrices and linear transformations. San Francisco, California, W. H. Freeman and Co. 1960.

10. Friedland, Bernard. Least squares filtering and prediction of nonstationary sampled data. Information and Control 1: 297-313. 1958.

11. Fulks, Watson. Advanced calculus. New York, N.Y., John Wiley and Sons, Inc. 1962.

12. Gibson, J. E. and Pearson, J. B., Jr. The state variable formulation in problems of automatic control. Unpublished dittoed paper presented at the Special Session on State Variables at the 1963 National Electronics Conference. Ames, Iowa, Department of Electrical Engineering, Iowa State University. ca. 1963.

13. Hague, D. S. Three-degree-of-freedom problem optimization formulation. U.S. Air Force Systems Command, Wright-Patterson Air Force Base, Ohio. Research and Technology Division Flight Dynamics Laboratory Technical Document Report FDL-TDR-64-1, Part 1, Vol. 3. 1964.

14. Hammon, Robert L. An application of random process theory to gyro drift analysis. Institute of Radio Engineers Transaction on Aeronautical and Navigational Electronics ANE-7: 84-91. 1960.

15. Hammon, Robert L. Effects on inertial guidance systems of random error sources. Institute of Radio Engineers Transactions on Aerospace and Navigational Electronics ANE-9: 215-230. 1962.

16. Kalman, R. E. New methods in Wiener filtering theory. In Bogdanoff, John L. and Kozin, Frank, eds. Proceedings of the First Symposium on Engineering Application of Random Function Theory and Probability. Pp. 270-388. New York, N.Y., John Wiley and Sons, Inc. 1963.

17. Koschmann, A. H. Time-varying filters for nonstationary signals on a finite interval. Institute of Radio Engineers Convention Record, Part 4, 3: 102-105. 1955.

18. Lampard, D. G. The response of linear networks to suddenly applied random noise. Institute of Radio Engineers Transactions on Circuit Theory CT-2: 49-57. 1955.

19. Laning, J. Halcombe, Jr. and Battin, Richard H. Random processes in automatic control. New York, N.Y., McGraw-Hill Book Co., Inc. 1956.

20. Lee, E. G. and Markus, L. Foundations of optimum control theory. Mimeographed manuscript used in Advanced Control Topics course. Minneapolis, Minn., Department of Electrical Engineering, University of Minnesota. 1965.

21. Miller, Kenneth S. An introduction to the calculus of finite differences and differential equations. New York, N.Y., Henry Holt and Co. 1960.

22. Mood, Alexander M. and Graybill, Franklin A. Introduction to the theory of statistics. 2nd ed. New York, N.Y., McGraw-Hill Book Co., Inc. 1963.

23. Parzen, Emanuel. Modern probability theory and its applications. New York, N.Y., John Wiley and Sons, Inc. 1960.

24. Parzen, Emanuel. Stochastic processes. San Francisco, California, Holden-Day, Inc. 1962.

25. Rand Corporation. A million random digits with 100,000 normal deviates. Glencoe, Ill., Glencoe, Ill. Free Press. 1955.

26. Sorenson, H. W. Kalman filtering techniques. Advances in Control Systems 3: 219-292. 1966.

27. Uhlenbeck, G. E. and Orstein, L. S. On the theory of Brownian motion. Physical Review 36: 823-841. 1930.

28. Zadeh, Lotfi A. and Desoer, Charles A. Linear system theory. New York, N.Y., McGraw-Hill Book Co., Inc. 1963.

VIII. ACKNOWLEDGMENTS

The author wishes to express his appreciation to the following people for their assistance in making this dissertation possible: Dr. Robert Grover Brown for his guidance and helpful suggestions during the author's studies at Iowa State University; Dr. David R. Glandorf, formerly with the Department of Aerospace Engineering at Iowa State University, for the use and explanation of his gradient computer program; and Mrs. Marjorie Wibholm, for typing the dissertation.

The author also wishes to express his appreciation to the A. C. Electronics Division of the General Motors Corporation of Milwaukee, Wisconsin, and in particular, Dr. Keith Struthers of that organization, for furnishing the gyro drift rate data analyzed in this study.

IX. APPENDIX A — METHOD OF STEEPEST DESCENT

A. Theory of Steepest Descent (20)

Suppose it is desirable to minimize a real valued function of m variables

$$C = C(u_1, u_2, \cdots, u_m) \tag{A.1}$$

in some domain (where C is continuous with continuous partial derivatives) by finding the values $u_1^*, u_2^*, \cdots, u_m^*$ which produce this minimum. These minimum values are a solution to the equation

$$\underline{\nabla} C = \begin{bmatrix} \frac{\partial C}{\partial u_1} \\ \vdots \\ \frac{\partial C}{\partial u_m} \end{bmatrix} = \underline{0} \tag{A.2}$$

This equation, however, is not always solvable in closed form. Thus, some numerical method must be used. One such method is the method of steepest descent.

As the name implies, this method seeks, from a given point $u_1^o, u_2^o, \cdots, u_m^o$, the steepest path to $u_1^*, u_2^*, \cdots, u_m^*$. It can be shown that the direction of this path is directed along the negative gradient vector $\underline{\nabla} C$.

Furthermore, the point $u_1^*, u_2^*, \cdots u_m^*$, is the solution of the differential equation

$$\frac{du_i(\sigma)}{d\sigma} = - k \frac{\partial C(u_1, u_2, \cdots, u_m)}{\partial u_i}, \quad i = 1, \cdots, m \tag{A.3}$$

as $\sigma \to \infty$, where $k > o$. This can be extended to a discrete process. That is, to move down the path of steepest descent, a step in the direction

of the negative gradient is taken. Thus, from a point $(u_1^{(i)}, \cdots, u_m^{(i)})$, a new point

$$u_j^{(i+1)} = u_j^{(i)} - k \frac{\partial C}{\partial u_j}(u_1^{(i)}, \cdots, u_m^{(i)}), \quad j = 1, \cdots, m \tag{A.4}$$

is computed, where k is a positive constant which determines the step size.

To achieve an absolute minimum, however, one must start from a point $(u_1^{(o)}, \cdots, u_m^{(o)})$ that is located in a neighborhood $N(u_1^*, \cdots, u_m^*)$ of the absolute minimum. Otherwise, the path of steepest descent may be one to a relative minimum or a saddle point. Usually, a satisfactory starting point can be estimated from a knowledge of the problem at hand.

B. Step-Size Control

Step-size control is basically a set of tests, mainly of a logical nature, which determine a suitable step-size, and hence a variable perturbation for each successive iteration of the steepest descent. This control is necessary in order to keep computer time within reasonable bounds, and to insure convergence to a minimum.

The step-size control used in this study is discussed in detail by Hague (13).

X. APPENDIX B — DISCRETE FILTERS

The output y of the linear time-invariant discrete filter H of Figure B.1 is related to the input u by the superposition equation

$$y(t) = \sum_{i=o}^{\infty} h(t - i)\, u(i), \quad t = 0, 1, 2, \cdots \tag{B.1}$$

The set of weighting coefficients h(t - i), which is sometimes called the one-sided Green's function (21), is defined by (10)

$$h(t - i) = \text{response of H to } \delta(t, i) \tag{B.2}$$

where

$$\delta(t - i) = \begin{cases} 1, & t = i \\ 0, & t \neq i \end{cases} \tag{B.3}$$

Since H is physically realizable,

$$h(t - i) = 0 \tag{B.4}$$

for $i > t$. Therefore, Equation B.1 can be written as

$$y(t) = \sum_{i=o}^{t} h(t - i)\, u(i), \quad t = 0, 1, 2, \cdots$$

$$= \sum_{i=o}^{t} h(i)\, u(t - i), \quad t = 0, 1, 2, \cdots$$

$$= h * u \tag{B.5}$$

B.5 is known as the convolution of h(t) and u(t).

A. Maclaurin Transform (4)

If we obtain a function F(x) by the Maclaurin expansion

$$F(x) = \sum_{t=o}^{\infty} f(t)x^t, \quad |x| < R \tag{B.6}$$

of a sequence $\{f(t)\}$, which will converge for some radius $|x| < R$,

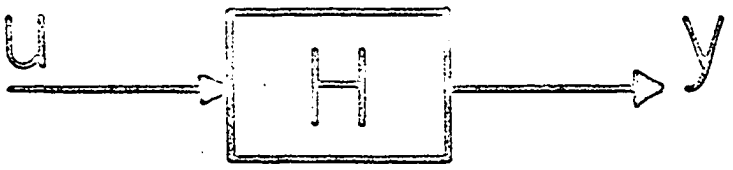

Figure B.1. Linear time-invariant discrete filter H

F(x) is the Maclaurin transformation of f(t) with an inverse transformation

$$f(t) = \frac{F^{(t)}(o)}{t!} \tag{B.7}$$

We shall write

$$F(x) = \mathcal{M}\{f(t)\},\ f(t) = \mathcal{M}^{-1}\{F(x)\} \tag{B.8}$$

as transform pairs.

Some useful transform pairs are

$$\mathcal{M}\{\delta(t, i)\} = x^i,\ |x| < \infty \tag{B.9}$$

$$\mathcal{M}\{\rho^t\} = (1 - \rho x)^{-1},\ |x| < \rho^{-1} \tag{B.10}$$

$$\mathcal{M}\{h * u\} = \mathcal{M}\{h(t)\} \cdot \mathcal{M}\{u(t)\} \tag{B.11}$$

and

$$\mathcal{M}\{f(t - i)\} = x^i F(x) \tag{B.12}$$

if $f(t) = 0$ for $t < o$.

B. Determination of Weighting Coefficients

We wish to determine the set of weighting coefficients corresponding to a difference equation. That is, if we have a difference equation of the type

$$\sum_{i=o}^{k} a_i y(t - i) = u(t) \tag{B.13}$$

we wish to determine the response of H of Figure B.1 to $\delta(t, i)$ that will yield a solution to Equation B.13 in the form of Equation B.5.

If we let $u(t) = \delta(t, o)$, $y(t)$ of Equation B.13 will be the response due to $\delta(t, o)$. That is, the solution of

$$\sum_{i=o}^{k} a_i y(t - i) = \delta(t, o) \tag{B.14}$$

will be $y(t) = h(t)$. This can be shown by operating on each side of Equation B.5 with the Maclaurin transform. We have

$$Y(x) = H(x) \cdot U(x) \tag{B.15}$$

or

$$Y(x) = H(x) \tag{B.16}$$

for $u(t) = \delta(t, o)$.

Operating on each side of Equation B.14, we have

$$\sum_{i=o}^{k} a_i x^i Y(x) = 1 \tag{B.17}$$

or

$$Y(x) = H(x) = \frac{1}{\sum_{i=o}^{k} a_i x^i} \tag{B.18}$$

The denominator of Equation B.18 can be factored. Thus

$$H(x) = \frac{1}{a_o}\left[\frac{1}{(1 - \rho_1 x)(1 - \rho_2 x) \cdots (1 - \rho_k x)}\right] \tag{B.19}$$

Assuming all the ρ_i's are unique, a partial fraction expansion yields

$$H(x) = \frac{c_1}{1 - \rho_1 x} + \frac{c_2}{1 - \rho_2 x} + \cdots + \frac{c_k}{1 - \rho_k x} \tag{B.20}$$

Using the transform pair[9] of Equation B.10, we have

$$h(t) = \sum_{i=1}^{k} c_i (\rho_i)^t, \quad t = 0, 1, 2, \cdots \tag{B.21}$$

Now,

$$(\rho_i)^t = e^{(\ln\rho_i)t}, \quad i = 0, 1, \cdots, k \tag{B.22}$$

If ρ_i is real negative,

$$(\rho_i)^t = e^{(\ln|\rho_i|t}\, e^{j\pi t} = e^{(\ln|\rho_i|)t} \cos\pi t \tag{B.23}$$

for integer values of t. If ρ_i and ρ_i^* are complex pairs,

$$\begin{aligned} c_i(\rho_i)^t + c_i^*(\rho_i^*)^t &= e^{\ln|c_i|}\, e^{j\arg c_i}\, e^{(\ln|\rho_i|)t}\, e^{j(\arg\rho_i)t} \\ &\quad + e^{\ln|c_i|}\, e^{-j\arg c_i}\, e^{(\ln|\rho_i|)t}\, e^{-j(\arg\rho_i)t} \\ &= 2|c_i| e^{(\ln|\rho_i|)t} \cos[(\arg\rho_i)t + \arg c_i] \end{aligned} \tag{B.24}$$

for integer values of t.

Thus, Equation B.21 may be written as

$$\begin{aligned} h(t) = &\sum_{\rho_i>0} c_i e^{(\ln\rho_i)t} + \sum_{\rho_i<0} c_i e^{(\ln|\rho_i|)t} \cos\pi t \\ &+ \sum_{\rho_i \text{ complex}} 2|c_i| e^{(\ln|\rho_i|)t} \cos[(\arg\rho_i)t + \arg c_i] \end{aligned} \tag{B.25}$$

XI. APPENDIX C — EXTENDING THE MODEL FOR NON-INTEGER VALUES OF TIME

As stated in Section D of Chapter III, Equations 3.24, 3.25 and 3.26 are not defined for a scale factor d_ξ of ξ that is not an integer multiple of the scale factor d_t of t. However, we can define a random process $\{\tilde{X}(\tau),\ \tau\epsilon\tilde{T}\}$, where $\tilde{T}$ is the set of positive real numbers, and where

$$\tilde{X}(t) = X(t) \tag{C.1}$$

for $t\epsilon T$. For other values of $\tau\epsilon\tilde{T}$, $\tilde{X}(\tau)$ is defined to be a reasonable interpolation of X(t). Then $\tilde{X}(\tau)$ can be represented by the vector difference equation

$$\underline{\tilde{X}}(\tau) = \Theta(\tau,\ \tau - \Delta\tau)\ \underline{\tilde{X}}(\tau - \Delta\tau) + \underline{W}(\tau,\ \tau - \Delta\tau) \tag{C.2}$$

for all $\tau\epsilon\tilde{T}$, and all $\Delta\tau\epsilon\tilde{T}$, where

$$\sum_{i=o}^{k} a_i\tilde{X}(t - i) = N(t) \tag{C.3}$$

for $t\epsilon T$.

One reasonable interpolation is to define powers of a matrix other than integer powers. One way to do this is by the Lagrange interpolating polynomial as described by Zadeh and Desoer (28). That is, if the minimal polynomial of matrix A has distinct roots, and if the function to be interpolated, $f(\lambda)$, is an analytic function in an open set containing the eigenvalues $\lambda_1, \cdots, \lambda_k$ of A, then

$$f(A) = \sum_{j=1}^{k} \left[\frac{f(\lambda_j)}{\prod_{\substack{i=1 \\ i \neq j}}^{k} (\lambda_j - \lambda_i)} \prod_{\substack{i=1 \\ i \neq j}}^{k} (A - \lambda_i I) \right] \tag{C.4}$$

In the investigations of this study, it is highly improbable that the eigenvalues would not be distinct, thus the minimal polynomial of A would have distinct roots.

In this case, the function is

$$f(\lambda) = \lambda^{\Delta\tau/d_t}, \quad \Delta\tau \epsilon \tilde{T} \tag{C.5}$$

which is analytic everywhere in the complex plane except for an axis extending from the origin to the point at infinity. This axis can be chosen anywhere so that it doesn't pass through any of the eigenvalues of A except for a possible eigenvalue at the origin, the occurrence of which is improbable. However, if an eigenvalue of A lies on the negative real axis, the function of Equation C.4 is a complex valued function, which is not permissible. We are justified, however, to discard a model of order k for which the transition matrix A has a negative eigenvalue. This eigenvalue indicates that the process being modeled has an oscillatory component, and the negative eigenvalue restricts the frequency of this oscillation to π/d_t radians per second. Thus, a model of order k + 1 is necessary to remove this restriction by allowing a pair of complex eigenvalues.

Thus, we may define the matrix A to a power as

$$A^{\Delta\tau/d_t} = \sum_{j=1}^{k} \left[\frac{\lambda_j^{\Delta\tau/d_t}}{\prod_{\substack{i=1 \\ i\neq j}}^{k} (\lambda_j - \lambda_i)} \prod_{\substack{i=1 \\ i\neq j}}^{k} (A - \lambda_i I) \right], \ \Delta\tau\epsilon\tilde{T} \tag{C.6}$$

and the transition matrix of Equation C.2 is

$$\Theta(\tau, \tau - \Delta\tau) = A^{\Delta\tau/d_t}, \ \Delta\tau\epsilon\tilde{T} \tag{C.7}$$

The problem of defining $\underline{W}(\tau, \tau - \Delta\tau)$ is not as simple as defining $\Theta(\tau, \tau - \Delta\tau)$. Let us first consider the case where $\Delta\tau$ is greater than d_t.

Case I: $\Delta\tau > d_t$

If $\Delta\tau$ is greater than d_t, as in Figure C.1, the input sequence of the equation

$$\tilde{\underline{X}}(\tau) = A\tilde{\underline{X}}(\tau - d_t) + \underline{b}N(\tau), \ \tau = d_t, 2d_t, \cdots \tag{C.8}$$

is added to the augmented system $[\frac{\Delta\tau}{d_t}]$ (greatest integer less than $\frac{\Delta\tau}{d_t}$) times with probability $1 + [\frac{\Delta\tau}{d_t}] - \frac{\Delta\tau}{d_t}$ and $[\frac{\Delta\tau}{d_t}] + 1$ times with probability $\frac{\Delta\tau}{d_t} - [\frac{\Delta\tau}{d_t}]$ during the time interval $\Delta\tau$ at times 1, 2, $\cdots$, $[\frac{\Delta\tau}{d_t}]$ or $[\frac{\Delta\tau}{d_t}] + 1$.

Equation C.2 then becomes

$$\tilde{\underline{X}}(\tau) = A^{\Delta\tau/d_t} \tilde{\underline{X}}(\tau - \Delta\tau) + \sum^{[\frac{\Delta\tau}{d_t}]-1} A^j \underline{b}N([\tfrac{\Delta\tau}{d_t}] - j), \ \tau = \Delta\tau, 2\Delta\tau, \cdots \tag{C.9}$$

with probability $1 + [\frac{\Delta\tau}{d_t}] - \frac{\Delta\tau}{d_t}$, and

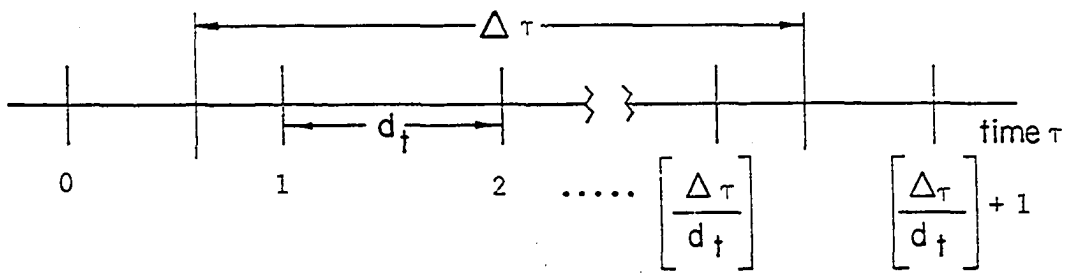

Figure C.1. Illustration of the time scale when $\Delta\tau > d_t$

$$\tilde{\underline{X}}(\tau) = A^{\Delta\tau/d_t}\,\tilde{\underline{X}}(\tau - \Delta\tau) + \sum_{j=o}^{[\frac{\Delta\tau}{d_t}]} A^j \underline{b} N([\tfrac{\Delta\tau}{d_t}] + 1 - j),$$

$$\tau = \Delta\tau,\ 2\Delta\tau,\ \cdots \qquad \text{(C.10)}$$

with probability $\frac{\Delta\tau}{d_t} - [\frac{\Delta\tau}{d_t}]$. Thus, the random sequence $\underline{W}(\tau, \tau - \Delta\tau)$ as a covariance matrix

$$K_{\underline{W}} = \sigma^2 \sum_{j=o}^{[\frac{\Delta\tau}{d_t}]-1} A^j\, \underline{b}\underline{b}'A'^j \qquad \text{(C.11)}$$

with probability $1 + [\frac{\Delta\tau}{d_t}] - \frac{\Delta\tau}{d_t}$, and

$$K_{\underline{W}} = \sigma^2 \sum_{j=o}^{[\frac{\Delta\tau}{d_t}]} A^j\, \underline{b}\underline{b}'A'^j \qquad \text{(C.12)}$$

with probability $\frac{\Delta\tau}{d_t} - [\frac{\Delta\tau}{d_t}]$.

We can use an average covariance matrix

$$\bar{K}_{\underline{W}} = \sigma^2 \sum_{j=o}^{[\frac{\Delta\tau}{d_t}]-1} A^j \underline{b}\underline{b}'A'^j + \sigma^2 (\tfrac{\Delta\tau}{d_t} - [\tfrac{\Delta\tau}{d_t}])\, A^{\frac{\Delta\tau}{d_t}}\, \underline{b}\underline{b}'A'^{\frac{\Delta\tau}{d_t}} \qquad \text{(C.13)}$$

or we can keep track of the time and use the covariances in Equations C.11 and C.12 during the applicable time intervals.

Case II: $d_t > \Delta\tau$

If $\Delta\tau$ is less than d_t, as in Figure C.2, an input sequence $\underline{M}(\tau)$ of the equation

$$\tilde{\underline{X}}(\tau) = A^{\Delta\tau/d_t}\,\tilde{\underline{X}}(\tau - \Delta\tau) + \underline{M}(\tau),\ \tau = \Delta\tau,\ 2\Delta\tau,\ \cdots \qquad \text{(C.14)}$$

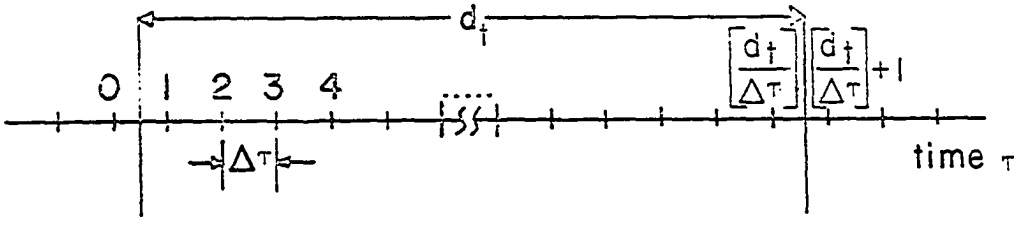

Figure C.2. Illustration of the time scale when $\Delta\tau < d_t$

is added to the augmented system $[\frac{d_t}{\Delta\tau}]$ times with probability $1 + [\frac{d_t}{\Delta\tau}] - \frac{d_t}{\Delta\tau}$ and $[\frac{d_t}{\Delta\tau}] + 1$ times with probability $\frac{d_t}{\Delta\tau} - [\frac{d_t}{\Delta\tau}]$ times during the time interval d_t, at times 1, 2, $\cdots$, $[\frac{d_t}{\Delta\tau}]$ or $[\frac{d_t}{\Delta\tau}] + 1$.

The equation

$$\underline{\tilde{X}}(\tau) = A\underline{\tilde{X}}(\tau - d_t) + \underline{b}N(\tau), \quad \tau = d_t, 2d_t, \cdots \tag{C.15}$$

becomes

$$\underline{\tilde{X}}(\tau) = A\underline{\tilde{X}}(\tau - d_t) + \sum_{j=0}^{[\frac{d_t}{\Delta\tau}]-1} A^{\frac{\Delta\tau}{d_t} j} M([\tfrac{d_t}{\Delta\tau}] - j), \quad \tau = d_t, 2d_t, \cdots \tag{C.16}$$

with probability $1 + [\frac{d_t}{\Delta\tau}] - \frac{d_t}{\Delta\tau}$, and

$$\underline{\tilde{X}}(\tau) = A\underline{\tilde{X}}(\tau - d_t) + \sum_{j=o}^{[\frac{d_t}{\Delta\tau}]} A^{\frac{\Delta\tau}{d_t} j} \underline{M}([\tfrac{d_t}{\Delta\tau}] + 1 - j), \quad \tau = d_t, 2d_t, \cdots \tag{C.17}$$

with probability $\frac{d_t}{\Delta\tau} - [\frac{d_t}{\Delta\tau}]$. Thus, the random sequence $\underline{W}(\tau, \tau - \Delta\tau) = \underline{M}(\tau)$ of Equation C.2 has a covariance matrix $K_{\underline{W}}$ which is the solution of the equation

$$\sum_{j=o}^{[\frac{d_t}{\Delta\tau}]-1} A^{\frac{\Delta\tau}{d_t} j} K_{\underline{W}} A'^{\frac{\Delta\tau}{d_t} j} = \sigma^2 \underline{b}\underline{b}' \tag{C.18}$$

with probability $1 + [\frac{d_t}{\Delta\tau}] - \frac{d_t}{\Delta\tau}$, and a solution of the equation

$$\sum_{j=o}^{[\frac{d_t}{\Delta\tau}]} A^{\frac{\Delta\tau}{d_t} j} \underline{K_W} A'^{\frac{\Delta\tau}{d_t} j} = \sigma^2 \underline{bb'} \qquad (C.19)$$

with probability $\frac{d_t}{\Delta\tau} - [\frac{d_t}{\Delta\tau}]$.

Equations C.18 and C.19 are solvable for $\underline{K_W}$ because there are, for a k^{th} order system, $\frac{k^2 + k}{2}$ independent linear equations of $\frac{k^2 + k}{2}$ unknowns. Once the solutions are found, we can use an average covariance matrix or a time-varying covariance matrix, as described for the case where $\Delta\tau$ is greater than d_t.