Optimal estimation of clock values and trends from finite data

Charles Greenhall
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, CA 91109 USA
E-mail: cgreenhall@jpl.nasa.gov

Abstract—We show how to solve two problems of optimal linear estimation from a finite set of phase data. Clock noise is modeled as a stochastic process with stationary $d$th increments. The covariance properties of such a process are contained in the generalized autocovariance function (GACV). We set up two principles for optimal estimation; these principles lead to a set of linear equations for the regression coefficients and some auxiliary parameters. The mean square errors of the estimators are easily calculated. The method can be used to check the results of other methods and to find good suboptimal estimators based on a small subset of the available data.

I. INTRODUCTION

Suppose that the phase residual $x(t)$ of a clock (or difference of clocks) is given at a finite set of times, $T = \{t_1, \ldots, t_n\}$, which don't have to be equally spaced. We consider two estimation targets: 1) the phase value $x(t_*)$ at some $t_* \notin T$; 2) the overall "trend" coefficient of $x(t)$. Depending on the model for $x(t)$ and what else is known, this can mean the long-term average value of phase or of its $d$th time derivative: frequency, drift rate, or aging rate ($d = 1, 2, \text{or } 3$). For each target, we want to calculate the linear estimator $\sum_{i=1}^{n} a_i x(t_i)$ that is optimal in the mean-square sense while satisfying an invariance condition that will be explained below.

To carry out this program, we need a stochastic model for $x(t)$. The general model used here is the class of stochastic processes with stationary $d$th increments, the subject of a monograph of Yaglom [1]. These include stationary processes, indefinite integrals of stationary processes, and all the power-law processes familiar to the time and frequency field: flicker PM, white FM, flicker FM, and so on. Yaglom showed how to solve problems of optimal prediction and filtering for these processes from their values on unbounded or bounded time intervals. For pure power-law processes, other authors have derived the mean square error (MSE) of the predictor that is based on the infinite past [2]–[4]; Boulanger and Douglas [5] calculated the MSE of two-point linear extrapolation. Vernotte et al. [6] calculated predictors and their MSEs based on extrapolation of least-squares linear and quadratic fits of equally spaced finite data sets. For finite-state clock models, a recursive optimal predictor and its MSE, based on all discrete-time past measurements, can be calculated from a Kalman filter [7]. The MSE of various suboptimal drift-rate estimators has been calculated for power-law noises [8]–[11]. Here we show how to calculate the regression coefficients $a_i$ and the MSE of the optimal linear estimators of both targets, using systems of linear equations that generalize the equations of orthogonal projection for stationary $x(t)$.

II. CLOCK NOISE MODELS

A real-valued, mean-square continuous1 stochastic process $x(t)$ is said to have stationary $d$th increments ($d \geq 1$) if for each $\tau$ the process

$$\Delta^d x(t) = \sum_{k=0}^{d} \binom{d}{k} (-1)^k x(t - k\tau)$$

is stationary. It is convenient to let $\text{SI}(d)$ denote the class of all such processes, and to let $\text{SI}(0)$ denote the stationary processes. Then $\text{SI}(d) \subset \text{SI}(d+1)$, and we define the degree of $x(t)$ as the least $d$ such that $x \in \text{SI}(d)$.

Everything we know about a process $x \in \text{SI}(d)$ is wrapped up in the $d$th increments (1), which do not change if we add a polynomial of degree $\leq d-1$ to $x(t)$. In this sense, a process in $\text{SI}(d)$ is ambiguous. Any use of these processes must take account of this ambiguity.

For any $x \in \text{SI}(d)$, Yaglom established a nonnegative spectral density function2 $S_x(f)$ that extends the notion of spectral density for stationary processes. Here, we are using the two-sided, even version: $S_x(-f) = S_x(f)$. For $d \geq 1$, $S_x(f)$ can diverge as $f \to 0$ but obeys the restrictions

$$\|f| \leq 1, \int_{|f| < 1} f^{2d} S_x(f) \, df < \infty, \quad \int_{|f| > 1} S_x(f) \, df < \infty. \tag{2}$$

The process also has an average trend coefficient, denoted by $c_d$, that can be defined as the infinite-time average of the stationary process $\Delta^d x(t)$:

$$c_d = \lim_{t_2 - t_1 \to \infty} \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \Delta^d x(t) \, dt,$$

the limit being taken in the mean-square sense. In this treatment, $c_d$ can be a random variable. We often want to get rid of $c_d$, and there are two ways to do it. First, if we know $c_d$, then we can consider the process $x_0(t) = x(t) - c_d t^d / d!$, which is also in $\text{SI}(d)$ but has trend coefficient zero. Second, if we don’t know $c_d$, then we can treat $x(t)$ as a member of $\text{SI}(d+1)$; as such, its trend coefficient $c_{d+1}$ is always zero. For example, a stationary process with an unknown mean can be treated as

---

1 This means that $E[x(t) - x(t)]^2 \to 0$ as $u \to t$.

2 Actually, a measure on the punctured frequency axis $f \neq 0$. 
Optimal estimation of clock values and trends from finite data

We show how to solve two problems of optimal linear estimation from a finite set of phase data. Clock noise is modeled as a stochastic process with stationary dth increments. The covariance properties of such a process are contained in the generalized autocovariance function (GACV). We set up two principles for optimal estimation; these principles lead to a set of linear equations for the regression coefficients and some auxiliary parameters. The mean square errors of the estimators are easily calculated. The method can be used to check the results of other methods and to find good suboptimal estimators based on a small subset of the available data.
a member of SI (1). If a random walk of phase (white FM), which is in SI (1), has an unknown slope (frequency) added to it, we can treat it as a member of SI (2). In this way we can get results that are invariant to the unknown trend.

Yaglom’s theory of SI (d) was based on the spectral density. For stationary processes, we also have the autocovariance (ACV) function

\[ s_x(t) = \int_{-\infty}^{\infty} e^{i2\pi ft} S_x(f) df, \]

which satisfies

\[ s_x(t-u) = E(x(t)x(u)). \]  

(4)

The estimation methods given here for SI (d) are based on a generalized ACV (GACV) function [12][13], also written as \( s_x(t) \), that can be obtained from \( S_x(f) \) and \( c_d \) by a generalized Fourier integral as follows:

\[
s_x(t) = \int_{|f| \leq 1} \left[ e^{i2\pi ft} - \sum_{k=0}^{2d-1} \frac{(i2\pi ft)^k}{k!} \right] S_x(f) df
+ \int_{|f| > 1} e^{i2\pi ft} S_x(f) df + (E c_d^2) \frac{(-t^2)^d}{(2d)!}.
\]

(5)

One may add a polynomial of degree \( \leq 2d-1 \) to \( s_x(t) \) without changing the value of any formula in which \( s_x(t) \) is properly used. With this ambiguity understood, Table I gives the GACV for power-law components of clock noise, specified by the one-sided spectral density of frequency, \( S_y^+(f) = 2(2\pi f)^2 S_y^-(f) \).

For any stationary process, the GACV is the ACV. The flicker PM entry is obtained by passing pure 1/f noise through a moving-average filter of width \( \tau \) to satisfy the second condition in (2). The trend coefficients are zero. The degree of a sum of noises is the maximum degree of the summands, and the GACV of the sum of orthogonal noises is the sum of the GACVs. By (5) we may and will assume that the GACV is an even function: \( s_x(-t) = s_x(t) \).

**TABLE I**

<table>
<thead>
<tr>
<th>Name</th>
<th>degree</th>
<th>( S_y^+(f) )</th>
<th>( s_x(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flicker PM (lowpass)</td>
<td>1</td>
<td>( h_1 f \frac{\sin(\pi f \tau)}{(\pi f \tau)^2} )</td>
<td>( -h_1 \frac{\tau^2}{4} \ln</td>
</tr>
<tr>
<td>White FM</td>
<td>1</td>
<td>( h_0 )</td>
<td>( -h_0 \frac{</td>
</tr>
<tr>
<td>Flicker FM</td>
<td>2</td>
<td>( h_{-1} f^{-1} )</td>
<td>( \frac{h_{-1} f^2 \ln</td>
</tr>
<tr>
<td>Random walk FM</td>
<td>2</td>
<td>( h_{-2} f^{-2} )</td>
<td>( \frac{-h_{-2} \pi^2</td>
</tr>
<tr>
<td>Flicker walk FM</td>
<td>3</td>
<td>( h_{-3} f^{-3} )</td>
<td>( \frac{-h_{-3} \pi^2</td>
</tr>
<tr>
<td>Random run FM</td>
<td>3</td>
<td>( h_{-4} f^{-4} )</td>
<td>( \frac{-h_{-4} \pi^4</td>
</tr>
</tbody>
</table>

Now we have to say how the GACV is used. In the following discussion, \( a(t) \) and \( b(t) \) denote functions that are zero except on some unspecified finite set of times; by using this notation we can freely perform and combine sums over \( t \) without worrying about the range of the summations. Such a function \( a(t) \) is said to satisfy the moment condition of order \( d \) if

\[ \sum_t a(t) t^j = 0, \quad j = 0, \ldots, d-1. \]  

(6)

In other words, if \( x(t) \) in \( \sum_t a(t) x(t) \) is replaced by a polynomial \( p(t) \) of degree \( \leq d-1 \), the result is zero. The coefficients of the \( d \)th increment \( (1) \) satisfy this condition. All the covariances needed for the estimation problems can be calculated by the following theorem.

**Theorem 1:** Let \( s_x(t) \) be the GACV of \( x(t) \), a process with stationary \( d \)th increments, where \( d \geq 1 \). If \( a(t) \) and \( b(t) \) satisfy the moment condition of order \( d \), then

\[
E \left( \sum_t a(t) x(t) \right) \left( \sum_t b(t) x(t) \right)
= \sum_{t,u} a(t) b(u) s_x(t-u).
\]

(7)

For stationary processes, (7) follows from (4), and no moment conditions are needed. For \( d \geq 1 \), we are only allowed to take the covariance of linear combinations of \( x(t) \) whose coefficients satisfy the moment condition. According to the theorem, we may do so as if (4) were true. In reality, \( s_x(t) \) is not an ACV, and (4) does not hold. Nevertheless, the entire formula (7) is correct, even though the corresponding terms of the expansions of its left and right sides are not equal.

We define \( M_d(x) \) as the set of random variables \( \sum_t a(t) x(t) \) whose coefficients \( a(t) \) satisfy the moment condition of order \( d \). Then \( M_d(x) \) is a linear subspace, and Theorem 1 tells us how to calculate the covariance of two members of \( M_d(x) \). It can be shown that any member of \( M_d(x) \) is a mean-square limit of linear combinations of \( d \)th increments of \( x(t) \); for this reason, if \( c_d = 0 \) then the members of \( M_d(x) \) have zero expectation. We also define \( M_d(x, T) \) as those members of \( M_d(x) \) whose coefficients are supported on the finite set \( T \).

**III. CLOCK PREDICTION**

Let \( x(t) \) have stationary \( d \)th increments. The estimation target is \( x(t_*) \), and the estimators are of form

\[
\hat{x}(t_*) = \sum_{t \in T} a(t) x(t),
\]

(8)

where \( T = \{ t_1, \ldots, t_n \} \) with \( n \geq d \). (Although this problem is called "prediction", there is no need to insist that \( t_* \geq t_i \). We now have to make the problem invariant to the ambiguity of \( x(t) \) with respect to polynomials \( p(t) \) of degree \( \leq d-1 \). We do so by insisting that the estimation error \( x(t_*) - \hat{x}(t_*) \) should not change if \( x(t) \) is replaced by \( x(t) + p(t) \). Thus the expression

\[
x(t_*) + p(t_*) - \sum_{t \in T} a(t) [x(t) + p(t)]
\]
should not depend on \(p(t)\). Consequently,
\[
\sum_{t \in T} a(t)p(t) = p(t_*) \tag{9}
\]
for any polynomial \(p(t)\) of degree \(\leq d - 1\); the estimator predicts \(p(t_*)\) perfectly from the values \(p(t), t \in T\). For \(d = 2\), if we add a constant phase and frequency to \(x(t)\), the estimate \(\hat{x}(t_*)\) should automatically adjust itself so that the error does not change. An equivalent statement of this invariance condition is that
\[
x(t_*) - \hat{x}(t_*) \in M_d(x), \tag{10}
\]
We may also express (9) as
\[
\sum_{t \in T} a(t)t^j = t_*^j, \quad j = 0, \ldots, d - 1. \tag{11}
\]
In matrix form, \(Ga = g\), where
\[
a = [a(t_1) \cdots a(t_n)]^T, \tag{12}
\]
\[
G = \begin{bmatrix}
1 & \cdots & 1 \\
t_1 & \cdots & t_n \\
\vdots & \vdots & \vdots \\
t_{d-1}^{d-1} & \cdots & t_n^{d-1}
\end{bmatrix}, \tag{13}
\]
\[
g = [1 \ t_* \ \cdots \ \hat{t}_*^{d-1}]^T. \tag{14}
\]
A random variable \(\hat{x}(t_*)\) of form (8) that satisfies the invariance condition is called a linear invariant estimator (LIE) of order \(d\) of \(x(t_*)\) based on \(x(t), t \in T\). Since \(n \geq d\), a LIE of order \(d\) exists, namely, the value at \(t_*\) of the interpolating polynomial determined by \(x(t_1), \ldots, x(t_d)\). For \(d = 2\), this LIE is the linear extrapolator of \(x(t_1)\) and \(x(t_2)\) to \(t_*\), which was treated by Boulanger and Douglas [5]. Some of the GACVs of Table 1 appear in their formulas. Supposing that \(t_1 < t_2 < t_*\), one can often optimize \(t_2 - t_1\) to give an MSE that is acceptably close to the minimum MSE that one can get from LIES based on the infinite past \(t \leq t_2\) [4]. The linear and quadratic extrapolators discussed by Vernotte et al. [6] are LIES of order 2 and 3, respectively. The ACVs tabulated in [6] are related to the GACVs used here.

Of all the order-\(d\) LIES of \(x(t_*)\), we want the one with the smallest MSE, called the best LIE (BLIE) of order \(d\). The set of LIES, call it \(\{\text{LIE}\}\), is determined by the inhomogeneous equations (11) for \(a(t)\). The difference of any two LIES is a random variable \(\sum_{t \in T} b(t) x(t)\) whose coefficients satisfy the corresponding homogeneous equations, that is, it belongs to the linear subspace \(M_d(x, T)\). Thus, \(\{\text{LIE}\}\) is an affine set, a shifted version of \(M_d(x, T)\) that does not pass through the origin. To find the closest point of \(\{\text{LIE}\}\) to \(x(t_*)\) in the mean-square sense, we drop a perpendicular from \(x(t_*)\) to \(\{\text{LIE}\}\). Thus, \(\hat{x}(t_*)\) has to be a member of \(\{\text{LIE}\}\) and also has to satisfy the orthogonality condition
\[
x(t_*) - \hat{x}(t_*) \perp M_d(x, T), \tag{15}
\]
which means that
\[
\mathbb{E}[x(t_*) - \hat{x}(t_*)]Y = 0 \quad \text{whenever} \ Y \in M_d(x, T). \tag{16}
\]

Figure 1 should make the geometry clear. It can be proved from basic facts about orthogonal projections that the BLIE \(\hat{x}(t_*)\) exists and is unique.

Let
\[
x(T) = [x(t_1) \cdots x(t_n)]^T, \quad b = [b_1 \cdots b_n]^T. \tag{17}
\]
Then (16) can be rewritten as
\[
\mathbb{E}[x(t_*) - a^Tx(T)] (x(T)^Ta) = 0 \quad \text{whenever} \ Gb = 0. \tag{18}
\]
Both factors on the left side of (18) belong to \(M_d(x)\); by Theorem 1, the expectation can be evaluated as if (4) were true. Doing so gives the condition
\[
(r^T - a^TR)b = 0 \quad \text{whenever} \ Gb = 0, \tag{19}
\]
where \(r\) is a column vector and \(R\) a symmetric \(n \times n\) matrix formed from the GACV of \(x(t)\):
\[
r = [s_x(t_1 - t_*), \ldots, s_x(t_n - t_*)]^T, \tag{20}
\]
\[
R = [s_x(t_i - t_j) : i, j = 1, \ldots, n]. \tag{21}
\]
In turn, condition (19) says that the row vector \(r^T - a^TR\) is orthogonal to all vectors \(b\) that are orthogonal to the rows of \(G\). Therefore, \(r^T - a^TR\) belongs to the row space of \(G\), that is, \(r - Ra = G^T\theta\) for some \(d\)-vector \(\theta = [\theta_0 \ \cdots \ \theta_{d-1}]^T\).

We now have the system of equations
\[
Ga = g, \quad Ra + G^T\theta = r, \tag{22}
\]
which constitute \(n + d\) equations in the \(n + d\) unknowns \(a, \theta\). They generalize the Yule-Walker equations
\[
Ra = r \tag{23}
\]
for stationary \(x(t)\); in that case, \(R\) is a genuine covariance matrix, and \(\hat{x}(t_*)\) is the orthogonal projection of \(x(t_*)\) on the unrestricted subspace generated by \(x(t), t \in T\). For \(d \geq 1\), \(R\) is symmetric if \(s_x(t)\) is even, but usually has both positive and negative eigenvalues.

After solving (22) for \(a\) and \(\theta\), we can calculate the mean square error of \(\hat{x}(t_*)\) as follows:
\[
\text{MSE} = \mathbb{E}[x(t_*) - \hat{x}(t_*)]^2
= \mathbb{E}[x(t_*) - a^Tx(T)] (x(t_*) - x(T)^Ta)
= s_x(0) - r^T a - a^T (r - Ra). \tag{379}
\]
by Theorem 1. Then, by (22),
\[
\text{MSE} = s_x(0) - r^T a - g^T \theta. \tag{24}
\]

We mention three methods for solving (22). First, if \( n = d \), then \( G \) is nonsingular, and we can solve (22) for \( a \) and then for \( \theta \). Second, if \( R \) is nonsingular, then we can write a solution in the following form (\( R^{-1} \) method):
\[
\Lambda = GR^{-1}G^T, \\
\theta = \Lambda^{-1}(GR^{-1}r - g), \\
a = R^{-1}(r - G^T\theta).
\]

Third, we can set up the system (22) as the single matrix equation
\[
\begin{bmatrix} R & G \\ G & 0 \end{bmatrix} \begin{bmatrix} a \\ \theta \end{bmatrix} = \begin{bmatrix} r \\ g \end{bmatrix}, \tag{25}
\]
and tell Matlab\(^4\) to solve it in one operation (“brute force” method). Although Matlab often says that the big matrix is badly scaled or nearly singular, the solution seems to work anyway in the cases the author has tried.

Before carrying out this solution, we should reduce the trend coefficient to zero as explained earlier, either subtracting the trend from \( x(t) \) if \( c_d \) is known, or increasing \( d \) by 1 if \( c_d \) is unknown. Then \( \tilde{x}(t_*) \) is unbiased for \( x(t_*) \) because \( x(t_*) - \tilde{x}(t_*) \in M_{d+1}(x) \), all of whose members have zero expectation. The penalty for increasing \( d \) is a greater MSE for the BLIE, because the set of LIEs shrinks as the order increases.

A. Examples

1. Model: white FM, \( h_0 = 1, d = 1 \), average frequency \( c_1 = 0 \), \( T = \{0, -1, \ldots, -10\} \), \( t_* = 5 \). We get the expected result: \( \hat{x}(5) = x(0) \), MSE = \( \frac{1}{2}h_0 t_* = 2.5 \).

2. The same example, but with an unknown frequency (phase slope) added. We calculate the BLIE of order 2 to get a predictor that is invariant to the added frequency. The result is the linear extrapolator \( \tilde{x}(5) = \frac{5}{2}x(0) - \frac{1}{2}x(-10) \), with MSE = 3.75. This is a simple demonstration of the penalty paid for a lack of knowledge of the trend. But, as \( T \) reaches farther and farther into the past, the estimator recovers the unknown frequency, and the MSE tends to \( \frac{1}{2}h_0 t_* \).

3. Model: white FM \( (h_0 = 2) + \) random walk FM \( (h_{-2} = 2.53 \times 10^{-5}), d = 2 \). Figure 2 shows normalized rms prediction error, \( \sqrt{\text{MSE}/\tau} \), of \( x(t + \tau) \) for three predictors: a) linear extrapolation from \( x(t), x(t - \tau) \), whose rms error is \( \sqrt{2\tau} \) times Allan deviation for \( \tau \); b) optimal prediction based on the entire discrete past \( x(t - 10n), n = 0, 1, \ldots, \), calculated in closed form by solving for the stationary covariance matrix of a Kalman filter [7]; c) optimal prediction based on \( x(t - 10n), n = 0, \ldots, 5 \), calculated by the method given here.

4. Model: flicker FM, \( h_{-1} = 1, d = 2 \). For \( t_* = 8 \), the squares in figure 3 show the BLIE regression coefficients for the 33-point prediction set \( T = \{-32, -31, \ldots, 0\} \). Because the coefficients are small off the set \( \{-32, -31, -1, 0\} \), we also try the latter set for prediction (filled circles). Figure 4 shows \( \sqrt{\text{MSE}/\tau} \) for both \( T \) sets and for \( t_* = \tau = 1, 2, 4, \ldots, 256 \). The lower horizontal line is for optimal prediction from the continuous past, \( t \leq 0 \). The upper horizontal line is for simple linear extrapolation from \( T = \{-\tau, 0\} \) as before. We observe only a small error penalty for using the 4-point set instead of the 33-point set. As \( \tau \) becomes large, the linear extrapolator becomes better than the other predictors, which see only 32 units into the past of this long-memory process.

IV. TREND ESTIMATION

Let \( x(t) \) have stationary \( \delta \)-th increments \( (d \geq 0) \), with an unknown trend coefficient \( c_d \). Then
\[
x(t) = c_d \frac{d^d}{dt^d} + x_0(t), \tag{26}
\]
where \( x_0(t) \) has trend coefficient zero. Again there will be invariance and orthogonality conditions to determine the

---

\(^3\)The author does not know any conditions for \( R \) to be nonsingular.

\(^4\)Copyright by The MathWorks, Inc.
to the trend

If and the estimator gives the right answer

\[ G \]

of degree \( \leq d \) exists, namely, the coefficient of \( d \)-th degree is added to \( x(t) \), we obtain

\[
c_{d} - \hat{c}_{d} = c_{d} \left[ 1 - \frac{1}{d!} \sum_{t \in T} a(t) t^{d} \right] - \sum_{t \in T} a(t) [x_{0}(t) + p(t)].
\]

If \( c_{d} - \hat{c}_{d} \) is invariant to \( c_{d} \) and \( p(t) \), then

\[
\sum_{t \in T} a(t) t^{j} = 0, \quad j = 0, \ldots, d - 1,
\]

\[
\sum_{t \in T} a(t) t^{d} = d!,
\]

and the estimator gives the right answer \( c_{d} \) if \( x(t) \) is a polynomial of degree \( \leq d \). A random variable of form (27) is said to be a LIE of \( c_{d} \) if (29) and (30) hold. We write this in the matrix form \( Ga = g \) again, but with different definitions of \( G \) and \( g \):

\[
G = \begin{bmatrix} 1 & \cdots & 1 \\ t_{1} & \cdots & t_{n} \\ \vdots & \vdots & \vdots \\ t_{1}^{d} & \cdots & t_{n}^{d} \end{bmatrix},
\]

\[
g = \begin{bmatrix} 0 \cdots 0 \ d! \end{bmatrix}^{T}.
\]

Assume that \( T \) has at least \( d + 1 \) points. Then a LIE of \( c_{d} \) exists, namely, the coefficient of \( t^{d}/d! \) in the interpolating polynomial determined by \( x(t_{1}), \ldots, x(t_{d+1}) \). Moreover, the difference of any two LIEs is in the subspace \( M_{d+1}(x, T) \), the set of \( \sum_{t \in T} b(t) x(t) \) such that \( b(t) \) satisfies the moment condition of order \( d + 1 \). As with the prediction problem, a LIE \( c_{d} \) is the BLIE if it satisfies the orthogonality condition

\[
c_{d} - \hat{c}_{d} = M_{d+1}(x, T).
\]

By (28), (29), and (30),

\[
c_{d} - \hat{c}_{d} = - \sum_{t \in T} a(t) x_{0}(t),
\]

which belongs to \( M_{d}(x_{0}) \) by (29). Also, \( M_{d+1}(x, T) = M_{d+1}(x_{0}, T) \subset M_{d}(x_{0}) \), because every member of \( M_{d+1}(x) \) has coefficients that kill the trend. Therefore, (33) can be written as

\[
E \left[ -a^{T} x_{0}(T) \right] \left[ x_{0}(T)^{T} b \right] = 0 \text{ whenever } Gb = 0,
\]

where \( x_{0}(T) \) is a column vector like \( x(T) \). Because both factors in the left side of (35) belong to \( M_{d}(x_{0}) \), we may use Theorem 1 for \( x_{0}(t) \) to evaluate this expression, giving the condition

\[
-a^{T} R b = 0 \text{ whenever } Gb = 0,
\]

which is like (22), except that now there are \( n + d + 1 \) equations and unknowns, the definitions of \( G \) and \( g \) are different, and \( r \) has become 0. The same solution methods are available.

By (33), \( c_{d} \) is unbiased for \( c_{d} \) because all the members of \( M_{d+1}(x) \) have mean zero. By (34), Theorem 1, and (37), we can calculate the MSE of \( \hat{c}_{d} \) by

\[
\text{MSE} = E \left[ (c_{d} - \hat{c}_{d})^{2} \right] = E \left[ -a^{T} x_{0}(T) \right] \left[ -x_{0}(T)^{T} a \right] - a^{T} Ra = -g^{T} \theta = -d! \theta_{d}.
\]

A. Examples

1. Model: \( x(t) = \text{white FM} + c_{1} t \) (unknown frequency), \( h_{0} = 1 \), \( d = 1 \), \( T = \{0, 1, \ldots, 10\} \). Result: \( \hat{c}_{1} = \frac{1}{10} [x(10) - x(0)] \), MSE = \( \frac{1}{10} \theta_{1}^{2} \).

2. Model: \( x(t) = x_{0}(t) + \frac{1}{2} c_{2} t^{2} \) where \( x_{0}(t) \) is white FM, flicker FM, or random walk FM, \( T = \{0, 1, \ldots, 10\} \). The regression coefficients for \( c_{2} \) are shown in Fig. 5. Even though white FM is in SI(1), we have to treat it as a member of SI(2) to extract a quadratic trend, independently of any linear trend that may also be present. For white FM, \( c_{2} \) is also the slope of the least-squares linear fit to the frequency data.
A. Prediction from equally spaced data

For fixed \( d \), it takes \( O(n^3) \) operations to solve the Yule-Walker equations (23) (when \( d = 0 \)) or their generalization (22) by general linear equation solving methods, where \( n \) is the number of elements in \( T \). If \( T \) is an equally spaced set of times, however, then \( R \) is a Toeplitz matrix. For stationary \( x(t) \), the Levinson-Durbin algorithm [15], which is a loop on \( n \), calculates the regression coefficients and the MSE in \( O(n^2) \) operations. The author has been able to extend this algorithm to the case \( d \geq 1 \) while keeping the \( O(n^2) \) property for fixed \( d \). In cases that have been tried, the results of the two general algorithms (\( R^{-1} \) and brute force) and the extended Levinson algorithm agree within roundoff error.

B. Trend from symmetric data

Suppose that \( T \) is symmetric about some point, which we assume to be zero; then \( -T = T \). It can be shown that the optimal trend coefficient estimator \( \hat{e}_d \) has coefficients that are even \( (a(t) = a(-t)) \) if \( d \) is even, and odd \( (a(t) = -a(-t)) \) if \( d \) is odd. In either case one can set up equations like (37) for \( a(t), t \geq 0 \), and a smaller auxiliary vector \( \theta \). Thus, the dimension \( n + d + 1 \) of the system (37) can be reduced by approximately a factor of two.

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES


