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 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; with the help of the GACV, 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 dth time derivative: frequency, drift rate, or aging rate ($d = 1, 2,$ 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 dth 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 based on 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]. For finite-state clock models, a recursive predictor and its MSE, based on all discrete-time past measurements, can be calculated from a Kalman filter [9]. The MSE of various suboptimal drift-rate estimators has been calculated for power-law noises [5]–[8]. 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 continuous[1] stochastic process $x(t)$ is said to have stationary dth increments ($d \geq 1$) if for each τ the process

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

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

Everything we know about a process $x \in \mathrm{SI}(d)$ is wrapped up in the dth 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 $\mathrm{SI}(d)$ is ambiguous. Any use of these processes must take account of this ambiguity.

For any $x \in \mathrm{SI}(d)$, Yaglom established a two-sided spectral density function[2] $S_x(f) \geq 0$. 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

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

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 $\Delta_1^d x(t)$:

$$c_d = \lim_{t_2 - t_1 \to \infty} \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \Delta_1^d x(t)\, dt,$$

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 $\mathrm{SI}(d)$ but has trend coefficient zero. Second, if we don't know c_d, then we may treat $x(t)$ as a member of $\mathrm{SI}(d+1)$; as such, its trend coefficient c_{d+1} is always zero. For example, a stationary process with an unknown mean can be treated as a member of $\mathrm{SI}(1)$. If a random walk of phase (white FM), which is in $\mathrm{SI}(1)$, has an unknown slope (frequency) added to it, we can treat it as a member of $\mathrm{SI}(2)$. In this way we can get results that are invariant to the unknown trend.

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

[2] Actually, a measure on the punctured frequency axis $f \neq 0$.

Yaglom's treatment of $\mathrm{SI}(d)$ was based on the spectral density. For stationary processes, we have the autocovariance (ACV) function

$$s_x(t) = \int_{-\infty}^{\infty} e^{i2\pi f t} S_x(f)\, df, \tag{3}$$

which satisfies

$$s_x(t-u) = \mathrm{E}\, x(t)\, x(u). \tag{4}$$

The estimation methods given here are based on a generalized ACV (GACV) function [10][11], also called $s_x(t)$, that can be obtained from $S_x(f)$ and c_d by a generalized Fourier integral as follows:

$$s_x(t) = \int_{|f|\le 1} \left[e^{i2\pi f t} - \sum_{k=0}^{2d-1} \frac{(i2\pi f t)^k}{k!} \right] S_x(f)\, df + \int_{|f|>1} e^{i2\pi f t} S_x(f)\, df + \left(\mathrm{E}\, c_d^2\right) \frac{\left(-t^2\right)^d}{(2d)!}. \tag{5}$$

One may add a polynomial of degree $\le 2d-1$ 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, $S_y^+(f) = 2(2\pi f)^2 S_x(f)$. For any stationary phase component, the GACV is the ACV. The flicker PM entry is obtained by passing pure $1/f$ noise through a moving-average filter of width τ 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

GENERALIZED AUTOCOVARIANCE OF POWER-LAW NOISES

Name	degree	$S_y^+(f)$	$s_x(t)$
Flicker PM (lowpass)	1	$h_1 f \frac{\sin^2(\pi f \tau)}{(\pi f \tau)^2}$	$\frac{h_1}{4\pi^2}\left(\frac{3}{2} - \ln \tau\right),\ t = 0$ $-\frac{h_1}{4\pi^2} \ln \lvert t \rvert,\ t \gg \tau$
White FM	1	h_0	$-\frac{h_0 \lvert t \rvert}{4}$
Flicker FM	2	$h_{-1} f^{-1}$	$\frac{h_{-1} t^2 \ln \lvert t \rvert}{2}$
Random walk FM	2	$h_{-2} f^{-2}$	$\frac{h_{-2} \pi^2 \lvert t \rvert^3}{6}$
Flicker walk FM	3	$h_{-3} f^{-3}$	$-\frac{h_{-3} \pi^2 t^4 \ln \lvert t \rvert}{6}$
Random run FM	3	$h_{-4} f^{-4}$	$-\frac{h_{-4} \pi^4 \lvert t \rvert^5}{30}$

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 for d [12] if

$$\sum_t a(t)\, t^j = 0, \quad j = 0, \ldots, d-1. \tag{6}$$

In other words, if $x(t)$ in $\sum_t a(t)\, x(t)$ is replaced by a polynomial $p(t)$ of degree $\le d-1$, the result is zero. The coefficients of the dth 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 dth increments, where $d \ge 1$. If $a(t)$ and $b(t)$ satisfy the moment condition for d, then

$$\mathrm{E}\left(\sum_t a(t)\, x(t)\right)\left(\sum_t b(t)\, x(t)\right) = \sum_{t,u} a(t)\, b(u)\, s_x(t-u). \tag{7}$$

For stationary processes, (7) follows from (4), and no moment conditions are needed. For $d \ge 1$, 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.

It is convenient to define $M_d(x)$ as the set of random variables $\sum_t a(t)\, x(t)$ whose coefficients $a(t)$ satisfy the moment condition for 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 dth 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 dth increments. The estimation target is $x(t_\star)$, and the estimators are of form

$$\hat{x}(t_\star) = \sum_{t \in T} a(t)\, x(t), \tag{8}$$

where $T = \{t_1, \ldots, t_n\}$ with $n \ge d$. (Although this problem is called "prediction", there is no need to insist that $t_\star > \max t_i$.). We now have to make the problem invariant to the ambiguity of $x(t)$ with respect to polynomials $p(t)$ of degree $\le d-1$. We do so by insisting that the estimation error $x(t_\star) - \hat{x}(t_\star)$ should not change if $x(t)$ is replaced by $x(t) + p(t)$. Then the estimator, if applied to $p(t)$ itself, predicts $p(t_\star)$ perfectly from the values $p(t), t \in T$. Take $d = 2$, for example: if we add a constant phase and frequency to $x(t)$, the estimate $\hat{x}(t_\star)$ should automatically adjust itself so that the error does not change. An equivalent statement of this *invariance condition* is that

$$x(t_\star) - \hat{x}(t_\star) \in M_d(x). \tag{9}$$

By (8), the moment condition for the left side of (9) says that

$$\sum_{t \in T} a(t)\, t^j = t_\star^j, \quad j = 0, \ldots, d-1. \tag{10}$$

In matrix form, $Ga = g$, where

$$a = \begin{bmatrix} a(t_1) & \cdots & a(t_n) \end{bmatrix}^{\mathrm{T}}, \tag{11}$$

$$G = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_n \\ \vdots & & \vdots \\ t_1^{d-1} & \cdots & t_n^{d-1} \end{bmatrix}, \tag{12}$$

$$g = \begin{bmatrix} 1 & t_* & \cdots & t_*^{d-1} \end{bmatrix}^{\mathrm{T}}. \tag{13}$$

A random variable $\hat{x}(t_*)$ that satisfies (9) is called a *linear invariant estimator* (LIE) of $x(t_*)$ based on $x(t), t \in T$. Since $n \geq d$, a LIE 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_*, formulas for whose MSE were given by Boulanger and Douglas [13] for power-law phase noises. 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 using the infinite past $t \leq t_2$ [4].

Of all the LIEs of $x(t_*)$, we want the one with the smallest MSE, called the best LIE (BLIE). The set of LIEs, call it $\{\mathrm{LIE}\}$, is determined by the inhomogeneous equations (10) 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 homogenous equations, that is, it belongs to the linear subspace $M_d(x, T)$. Thus, $\{\mathrm{LIE}\}$ 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 $\{\mathrm{LIE}\}$ to $x(t_*)$ in the mean-square sense, we drop a perpendicular from $x(t_*)$ to $\{\mathrm{LIE}\}$. Thus, $\hat{x}(t_*)$ has to be a member of $\{\mathrm{LIE}\}$ and also has to satisfy the *orthogonality condition*

$$x(t_*) - \hat{x}(t_*) \perp M_d(x, T), \tag{14}$$

which means that

$$\mathrm{E}\left[x(t_*) - \hat{x}(t_*)\right] Y = 0 \text{ whenever } Y \in M_d(x, T). \tag{15}$$

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.

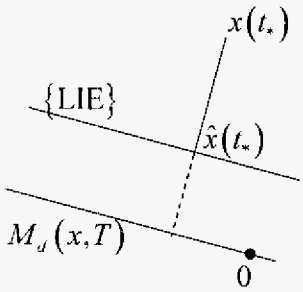


Fig. 1. Geometric interpretation of the conditions for a BLIE

Let

$$x(T) = \begin{bmatrix} x(t_1) & \cdots & x(t_n) \end{bmatrix}^{\mathrm{T}}, \; b = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^{\mathrm{T}}. \tag{16}$$

Then (15) can be rewritten as

$$\mathrm{E}\left[x(t_*) - a^{\mathrm{T}} x(T)\right] \left[x(T)^{\mathrm{T}} b\right] = 0 \text{ whenever } b^{\mathrm{T}} A = 0. \tag{17}$$

Both factors on the left side of (17) belong to $M_d(x)$; by Theorem 1, the expectation can be evaluated as if (4) were true. Doing so gives the condition

$$\left(r^{\mathrm{T}} - a^{\mathrm{T}} R\right) b = 0 \text{ whenever } Gb = 0, \tag{18}$$

where r is a column vector and R a symmetric $n \times n$ matrix formed from the GACV of $x(t)$:

$$r = \begin{bmatrix} s_x(t_1 - t_*) & \cdots & s_x(t_n - t_*) \end{bmatrix}^{\mathrm{T}}, \tag{19}$$

$$R = \left[s_x(t_i - t_j) : i, j = 1, \ldots, n\right]. \tag{20}$$

In turn, condition (18) says that the row vector $r^{\mathrm{T}} - a^{\mathrm{T}} R$ is orthogonal to all vectors b that are orthogonal to the rows of G. Therefore, $r^{\mathrm{T}} - a^{\mathrm{T}} R$ belongs to the row space of G, that is, $r - Ra = G^{\mathrm{T}} \theta$ for some d-vector $\theta = \begin{bmatrix} \theta_0 & \cdots & \theta_{d-1} \end{bmatrix}^{\mathrm{T}}$.

We now have the system of equations

$$Ga = g, \quad Ra + G^{\mathrm{T}} \theta = r, \tag{21}$$

which constitute $n + d$ equations in the $n + d$ unknowns a, θ. They generalize the Yule-Walker equations

$$Ra = r \tag{22}$$

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 $d \geq 1$, V is symmetric if $s_x(t)$ is even, but usually has positive and negative eigenvalues.

After solving (21) for a and θ, we can calculate the mean square error of $\hat{x}(t_*)$ as follows:

$$\begin{aligned} \mathrm{MSE} &= \mathrm{E}\left[x(t_*) - \hat{x}(t_*)\right]^2 \\ &= \mathrm{E}\left[x(t_*) - a^{\mathrm{T}} x(T)\right] \left[x(t_*) - x(T)^{\mathrm{T}} a\right] \\ &= s_x(0) - r^{\mathrm{T}} a - a^{\mathrm{T}}(r - Ra) \quad \text{by Theorem 1,} \\ \mathrm{MSE} &= s_x(0) - r^{\mathrm{T}} a - g^{\mathrm{T}} \theta \quad \text{by (21).} \end{aligned} \tag{23}$$

We mention three methods for solving (21). First, if $n = d$, then G is a nonsingular square matrix, and we can solve (21) for a and then for θ. Second, if R is nonsingular[3], then we can write a solution in the following form (R^{-1} method):

$$\begin{aligned} \Lambda &= G V^{-1} G^{\mathrm{T}}, \\ \theta &= \Lambda^{-1}\left(G R^{-1} r - g\right), \\ a &= R^{-1}\left(r - G^{\mathrm{T}} \theta\right). \end{aligned}$$

Third, we can set up the system (21) as the single matrix equation

$$\begin{bmatrix} R & G^{\mathrm{T}} \\ G & 0 \end{bmatrix} \begin{bmatrix} a \\ \theta \end{bmatrix} = \begin{bmatrix} r \\ g \end{bmatrix}, \tag{24}$$

and tell Matlab[4] to solve it in one operation ("brute force" method). Although Matlab often says that the big matrix is

[3]The author does not know any conditions for R to be nonsingular.

[4]Copyright by The MathWorks, Inc.

badly scaled or nearly singular, the solution seems to work anyway.

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 $\hat{x}(t_*)$ is unbiased for $x(t_*)$ because $x(t_*) - \hat{x}(t_*) \in M_d(x)$, 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.

A. Examples

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

2. The same example, but with an unknown frequency (phase slope) added. We perform the estimation with $d = 2$ to get an estimator that is invariant to the added frequency. The result is $\hat{x}(5) = \frac{3}{2}x(0) - \frac{1}{2}x(-10)$, $\mathrm{MSE} = 3.75$. 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, $\sqrt{\mathrm{MSE}}/\tau$, 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 τ; 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 [9]; c) optimal prediction based on $x(t-10n)$, $n = 0, \ldots, 5$, calculated by the method given here.

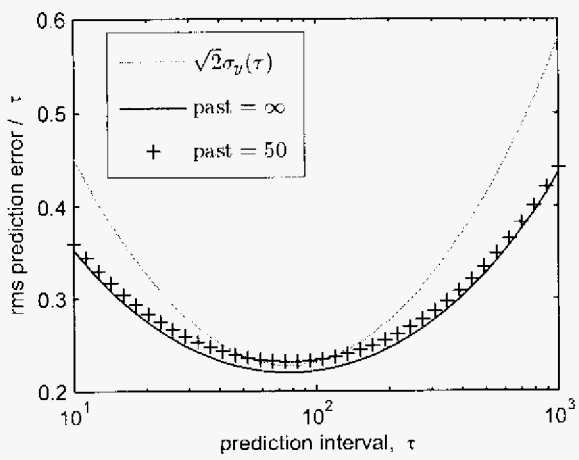


Fig. 2. Normalized rms error of predictors of a sum of white FM and random walk FM.

4. Model: flicker FM, $d = 2$, $h_{-1} = 1$. For $t_* = 8$, the squares in figure 3 show the regression coefficients for the 33-point prediction set $T = \{-32, -31, ..., 0\}$. 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 $\sqrt{\mathrm{MSE}}/\tau$ for both T sets and for $t_* = \tau = 1, 2, 4, \ldots, 256$. The lower horizontal line is $\sqrt{\mathrm{MSE}}/\tau$ for optimal prediction from the continuous past, $t \leq 0$. The upper horizontal line is $\sqrt{\mathrm{MSE}}/\tau$ 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 τ becomes large, the linear extrapolator becomes better than the other predictors, which see only 32 units into the past of this long-memory process.

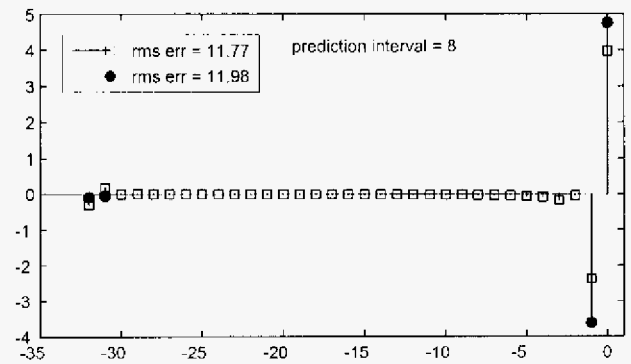


Fig. 3. Regression coefficients for predicting flicker FM

IV. Trend Estimation

Let $x(t)$ have stationary dth increments ($d \geq 0$), with an unknown trend coefficient c_d. Then

$$x(t) = c_d \frac{t^d}{d!} + x_0(t), \tag{25}$$

where $x_0(t)$ has trend coefficient zero. It is convenient to denote the GACV of $x_0(t)$ by $s_x(t)$ instead of $s_{x_0}(t)$. Again

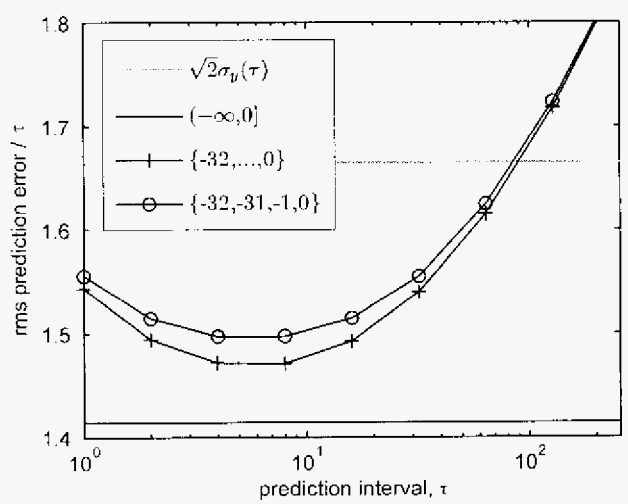


Fig. 4. Normalized rms errors of predictors of flicker FM

there will be invariance and orthogonality conditions to determine the optimal linear estimator $\hat{c}_d = \sum_{t\in T} a(t)\, x(t)$ of c_d from $x(t), t \in T$. The invariance condition is determined by the requirement that the estimation error $c_d - \hat{c}_d$ be unchanged if a polynomial of degree $\leq d$ is added to $x(t)$. Thus the error is invariant to the trend $c_d t^d/d!$ itself as well as to polynomials $p(t)$ of degree $\leq d-1$ added to $x(t)$. After rewriting (25) as $x(t) = c_d t^d/d! + x_0(t) + p(t)$, we obtain

$$c_d - \hat{c}_d = c_d \left[1 - \frac{1}{d!}\sum_{t\in T} a(t)\, t^d\right] - \sum_{t\in T} a(t)\left[x_0(t) + p(t)\right]. \tag{26}$$

If $c_d - \hat{c}_d$ is invariant to c_d and $p(t)$, then

$$\sum_{t\in T} a(t)\, t^j = 0, \quad j = 0, \ldots, d-1, \tag{27}$$

$$\sum_{t\in T} a(t)\, t^d = d!, \tag{28}$$

and the estimator gives the right answer c_d if $x(t)$ is a polynomial of degree $\leq d$. A random variable $\sum_{t\in T} a(t)\, x(t)$ is said to be a LIE of c_d if (27) and (28) hold. We write this in the matrix form $Ga = g$ again, but with different definitions of G and g:

$$G = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_n \\ \vdots & & \vdots \\ t_1^d & \cdots & t_n^d \end{bmatrix}, \tag{29}$$

$$g = \begin{bmatrix} 0 & \cdots & 0 & d! \end{bmatrix}^{\mathrm{T}}. \tag{30}$$

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 for $d+1$. As with the prediction problem, a LIE $\hat{c}_d = \sum_{t\in T} a(t)\, x(t)$ of c_d is the BLIE if it satisfies the orthogonality condition

$$c_d - \hat{c}_d \perp M_{d+1}(x, T). \tag{31}$$

By (26), (27), and (28),

$$c_d - \hat{c}_d = -\sum_{t\in T} a(t)\, x_0(t), \tag{32}$$

which belongs to $M_d(x_0)$ by (27). Also, $M_{d+1}(x, T) = M_{d+1}(x_0, T) \subset M_d(x_0)$, because every member of M_{d+1} has coefficients that kill the trend. Therefore, (31) can be written as

$$\mathrm{E}\left[-a^{\mathrm{T}} x_0(T)\right]\left[x_0(T)^{\mathrm{T}} b\right] = 0 \text{ whenever } Gb = 0, \tag{33}$$

where $x_0(T)$ is a column vector like $x(T)$. Because both factors in the left side of (33) belong to $M_d(x_0)$, we may use Theorem 1 for $x_0(t)$ to evaluate this expression, giving the condition

$$-a^{\mathrm{T}} Rb = 0 \text{ whenever } Gb = 0, \tag{34}$$

where R is defined by (20). By the same argument as before, $-Ra = G^{\mathrm{T}}\theta$ for some $(d+1)$-vector $\theta = \begin{bmatrix}\theta_0 & \cdots & \theta_d\end{bmatrix}^{\mathrm{T}}$. We arrive at the system of equations

$$Ga = g, \quad Ra + G^{\mathrm{T}}\theta = 0, \tag{35}$$

which is like (21), 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 (31), $\hat{c}_d$ is unbiased for c_d because all the members of $M_{d+1}(x)$ have mean zero. By (32), Theorem 1, and (35), we can calculate the MSE of $\hat{c}_d$ by

$$\begin{aligned} \mathrm{MSE} = \mathrm{E}\left(c_d - \hat{c}_d\right)^2 &= \mathrm{E}\left[-a^{\mathrm{T}} x_0(T)\right]\left[-x_0(T)^{\mathrm{T}} a\right] \\ &= a^{\mathrm{T}} Ra = -g^{\mathrm{T}}\theta = -d!\theta_d. \end{aligned} \tag{36}$$

A. Examples

1. Model: $x(t)$ = white FM + $c_1 t$ (unknown frequency), $d = 1$, $h_0 = 1$, $T = \{0, 1, ..., 10\}$. Result: $\hat{c}_1 = \frac{1}{10}\left[x(10) - x(0)\right]$, MSE = $\frac{1}{20}$.

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, $T = \{0, 1, ..., 10\}$. The regression coefficients for $\hat{c}_2$ 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.

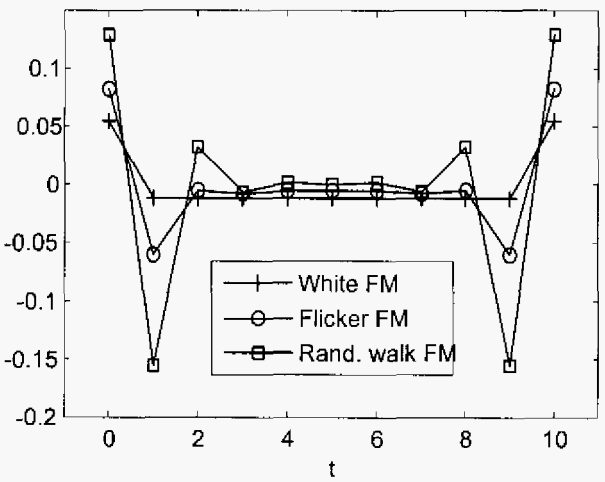


Fig. 5. Regression coefficients for estimating quadratic drift rate in the presence of white FM, flicker FM, or random walk FM.

V. Additional Results

A. Prediction from equally spaced data

Solving the prediction problem from the Yule-Walker equations (22) (when $d = 0$) or their generalization (21) by general linear equation solving methods requires $O(n^3)$ operations

for fixed d, 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 [14], 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 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 $\hat{c}_d$ 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 (35) for $a(t)$, $t \geq 0$, and a smaller auxiliary vector θ. Thus, the dimension $n + d + 1$ of the system (35) 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

[1] A. M. Yaglom, "Correlation theory of processes with random stationary nth increments," *American Mathematical Society Translations*, series 2, vol 8, pp 87–141, 1958.

[2] J. A. Barnes and D. W. Allan, "An approach to the prediction of coordinated universal time," *Frequency*, pp 1–8, Nov–Dec 1967.

[3] P. Tavella and M. Gotta, "Uncertainty and prediction of clock errors in space and ground applications," *14th European Frequency and Time Forum*, pp 77–81, Torino, 2000.

[4] L.-G. Bernier, "Use of the Allan deviation and linear prediction for the determination of the uncertainty on time calibrations against predicted timescales," *IEEE Trans. Instrum. Meas.*, vol 52, no 2, pp. 483–486, April 2003.

[5] M. A. Weiss and C. Hackman, "Confidence on the three-point estimator of frequency drift," *24th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp 451–459, 1992.

[6] V. A. Logachev and G. P. Pashev, Estimation of linear frequency drift coefficient of frequency standards, *1996 IEEE International Frequency Control Symposium*, pp. 960–963.

[7] G. Wei, Estimations of frequency and its drift rate, *IEEE Trans. Instrum. Meas.*, vol 46, no 1, pp 79–82, Feb 1997.

[8] C. A. Greenhall, "A frequency-drift estimator and its removal from modified Allan variance," *1997 IEEE International Frequency Control Symposium*, pp 358–365.

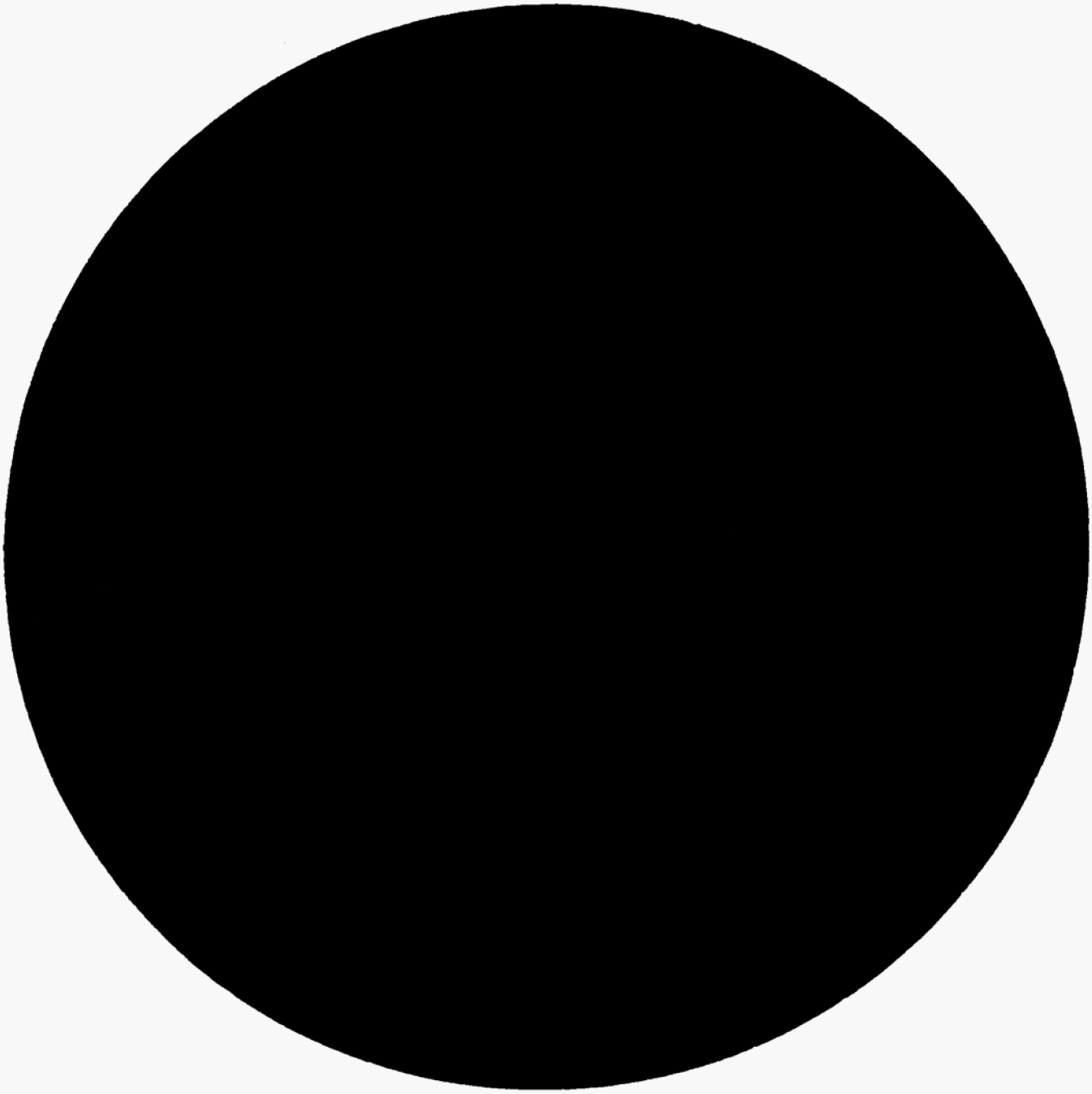
[9] J. A. Davis, C. A. Greenhall, and R. Boudjemaa, "The development of a Kalman filter clock predictor," *19th European Frequency and Time Forum*, Besançon, 2005.

[10] C. A. Greenhall, "A structure function representation theorem with applications to frequency stability estimation," *IEEE Trans. Instrum. Meas.*, vol IM-32, pp 364–370, 1983.

[11] C. A. Greenhall, "The generalized autocovariance: a tool for clock noise statistics," *The Tracking and Mission Operations Progress Report 42-137*, Jet Propulsion Laboratory, 1999, http://tmo.jpl.nasa.gov/tmo/progress_report/42-137/137J.pdf

[12] F. Vernotte, "Application of the moment condition to noise simulation and to stability analysis," *IEEE Trans. UFFC*, vol 49, no 4, pp 508–513, April 2002.

[13] J.-S. Boulanger and R. J. Douglas, "Traceability of frequency control systems," *1997 IEEE International Frequency Control Symposium*, pp 366–372, Orlando, 1997.

[14] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods, Second Edition*, §5.2, Springer, 1991.

End of File