Optimal prediction of clocks from finite data

V3

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

To be given at FPSAC '05 in honor of Adriano Garsia's 75th birthday

Abstract

This talk is about optimal linear prediction of processes with stationary dth increments, which serve as a class of models for random clock disturbances. The predictor is obtained by orthogonal projection on the affine space of estimators whose errors are invariant to additive polynomials of degree $< d$. The projection conditions give a system of linear equations that can be solved straightforwardly for the regression coefficients. If the data are equally spaced, then the predictor can be obtained by an extension of Levinson's algorithm.

1 Background

A clock can be evaluated only against other clocks. Therefore, the title actually refers to the prediction of the difference of two clocks. Each national timing center wants to keep its reference clock close to Coordinated Universal Time (UTC), or at least to keep track of the difference. UTC is a timescale, a virtual clock that is defined by its offsets from physical clocks. The offsets of UTC from the national reference clocks in units of nanoseconds, as calculated for a given month by the International Bureau of Weights and Measures from satellite observations, are disseminated around the middle of the following month. Because of this monthly schedule and delay, a timing center needs to be able to predict its offset up to 50 days from the last available data.

To predict a clock offset (also called "phase") and to evaluate the prediction uncertainty, a stochastic model for the clock's random fluctuations is needed. We observe samples of phase, but most of the underlying stationary noise mechanisms are buried one to three time derivatives deep. Accordingly, a suitable general model for clock phase is the class of stochastic processes with stationary dth increments, where $d = 1$, 2, or 3. Kolmogorov developed a theory for $d = 1$ in the 1940s and applied it to the study of turbulence. Theories for $d \geq 1$ were developed by A. M. Yaglom and M. S. Pinsker in the 1950s; Yaglom's version was published as an AMS translation paper [1]. Here I will give some of the elements of this theory, as filtered through my own preferences.

Let me quickly review wide-sense stationary processes. A real-valued process $X(t)$, continuous in mean-square, is stationary if

$$\mathrm{E}X(t) = c, \quad \mathrm{E}X(t)X(u) = s_X(t-u)$$

for some constant c and autocovariance function (ACV) $s_X(t)$. Then

$$s_X(t) = c^2 + \int_{-\infty}^{\infty} e^{i2\pi ft} S_X(f)\, df \tag{1}$$

where $S_X(f)\, df$ represents a finite Baire measure, even about 0, called the spectrum, but I am writing it as a spectral density. It could have an atom at $f = 0$, which gives rise to a random

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

constant component, but I assume that that is not the case. If $a(t)$ is a real-valued function of finite support on $\mathbb{R}$ (zero outside a finite set of reals), write

$$\tilde{a}(f) = \sum_t a(t) e^{i2\pi ft} .$$

For any $a(t)$ and $b(t)$ of finite support,

$$\begin{aligned} \mathrm{E}\sum_t a(t) X(t) &= c\sum_t a(t), & (2) \\ \mathrm{E}\left(\sum_t a(t) X(t)\right)\left(\sum_t b(t) X(t)\right) &= \sum_{t,u} a(t) b(u) s_X(t-u) & (3) \\ &= c^2 \left(\sum a(t)\right)\left(\sum b(t)\right) + \int_{-\infty}^{\infty} \tilde{a}(f) \overline{\tilde{b}(f)} S_X(f)\, df. & (4) \end{aligned}$$

We say that a real mean-square continuous process $X(t)$ has stationary dth increments $(d \geq 1)$, and write $X \in \mathrm{SI}(d)$, if for any τ the increment process

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

is stationary. Then $\mathrm{SI}(0)$ (stationary processes) $\subset \mathrm{SI}(d) \subset \mathrm{SI}(d+1)$.

Let us say that a function $a(t)$ of finite support kills $\mathcal{P}^{d-1}$ (polynomials of degree $\leq d-1$) if $\sum_t a(t) t^j = 0$ for $j = 0$ to $d-1$. Then $\tilde{a}(f)$ has a zero of order $\geq d$ at $f = 0$. For example, $a(t - k\tau) = \binom{d}{k}(-1)^k$ for $k = 0$ to d (t fixed) and 0 elsewhere, $\tilde{a}(f) = e^{i2\pi ft}\left(1 - e^{-i2\pi f\tau}\right)^k$.

Let $X \in \mathrm{SI}(d)$ and define $\mathcal{S}_d$ as the space of linear combinations $\sum_t a(t) X(t)$ where $a(t)$ kills $\mathcal{P}^{d-1}$. Then $\Delta_\tau^d X(t) \in \mathcal{S}_d$, and $\mathcal{S}_d$ is a non-closed subspace of the L^2 closure of the span of $\Delta_\tau^d X(t)$ for all τ, t. If T is a finite set of reals, let $\mathcal{S}_d(T)$ be the members of $\mathcal{S}_d$ with coefficient functions $a(t)$ supported on T.

For $X \in \mathrm{SI}(d)$, Yaglom established extensions of the mean and spectrum of a stationary process. There is a constant c_d, which I call the trend coefficient, such that

$$\mathrm{E}\sum_t a(t) X(t) = c_d \sum_t a(t) \frac{t^d}{d!} \tag{5}$$

for any $a(t)$ that kills $\mathcal{P}^{d-1}$. In fact, $c_d = \mathrm{E}\Delta_1^d X(t)$. The spectrum $S_X(f)\, df$ is now a measure on the punctured real line $\mathbb{R} \setminus \{0\}$, even with respect to 0, such that

$$\int_0^1 f^{2d} S_X(f)\, df < \infty, \quad \int_1^\infty S_X(f)\, df < \infty, \tag{6}$$

$$\mathrm{E}\left(\sum_t a(t) X(t)\right)\left(\sum_t b(t) X(t)\right) = c_d^2 \left(\sum a(t) \frac{t^d}{d!}\right)\left(\sum b(t) \frac{t^d}{d!}\right) + \int_{-\infty}^{\infty} \tilde{a}(f) \overline{\tilde{b}(f)} S_X(f)\, df. \tag{7}$$

whenever $a(t)$ and $b(t)$ kill $\mathcal{P}^{d-1}$. I am assuming that $X(t)$ contains no random trend component $Wt^d/d!$ where $EW = 0$. For any even measure $S_X(f)\, df$ satisfying (6) there is an $X \in \mathrm{SI}(d)$ whose $\mathcal{S}_d$ is a system of jointly Gaussian random variables.

Let $X \in \mathrm{SI}(d)$. If c_d is known, it can be eliminated by subtracting $c_d t^d/d!$ from $X(t)$. We can also treat X as a member of $\mathrm{SI}(d+1)$; then $S_X(f)$ stays the same but $c_{d+1} = 0$. Any indefinite integral $Y(t)$ of $X(t)$ is in $\mathrm{SI}(d+1)$, c_{d+1} for Y equals c_d for X, and $S_Y(f) = S_X(f) / (2\pi f)^2$. But there are members of $\mathrm{SI}(d+1)$ that cannot be obtained in this way.

My contribution to this theory was to establish an analog of the ACV [2][3]. For $X \in \mathrm{SI}(d)$, define the generalized autocovariance (GACV) function $s_X(t)$ by the following Fourier integral:

$$s_X(t) = c_d^2 \frac{(-t^2)^d}{(2d)!} + \left\{ \int_{|f|\le 1} \left[e^{i2\pi ft} - \sum_{k=0}^{2d-1} \frac{(i2\pi ft)^k}{k!} \right] + \int_{|f|>1} e^{i2\pi ft} \right\} S_X(f)\, df \bmod \mathcal{P}^{2d-1}. \quad (8)$$

For the GACV we may use any real-valued function that differs from the indicated formula by a polynomial of degree $< 2d$. In particular, the breakpoint 1 is arbitrary. Any GACV is continuous, and we may specify that $s_X(-t) = s_X(t)$. From (7), (3) follows easily, but only for $a(t)$ and $b(t)$ that kill $\mathcal{P}^{d-1}$. For example, if $d = 1$ then

$$\mathrm{E}[X(t+\tau) - X(t)]^2 = -s_X(\tau) + 2s_X(0) - s_X(-\tau) \quad (9)$$

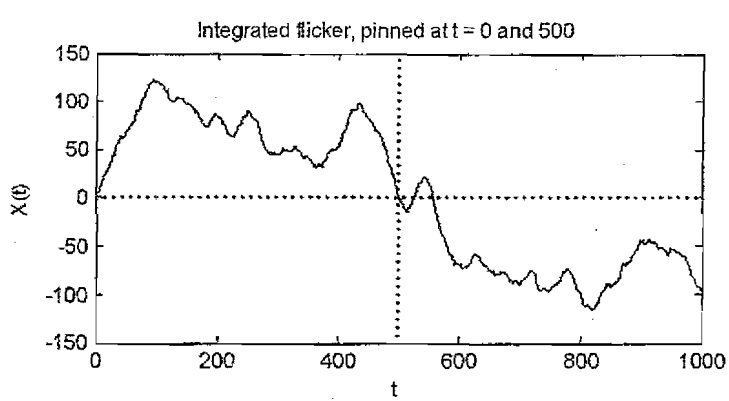
(the "structure function"). Once we have found $s_X(t)$, we can compute covariances of random variables in $\mathcal{S}_d$ by finite sums. But the GACV is not an ACV; we cannot compute $\mathrm{E}X(t)X(u)$ without more information.

Because of the mod $\mathcal{P}^{2d-1}$ arbitrariness of the GACV, we can calculate alternate versions by means of Fourier transforms in the complex t-domain. Here are some examples for "power-law" processes.

Name	d	$S_X(f)$	$s_X(t)$
Wiener process	1	$(2\pi f)^{-2}$	$-\frac{\lvert t\rvert}{2}$
Fractional Brownian motion [4]	1	$\lvert 2\pi f\rvert^{-2\delta}$, $\frac{1}{2} < \delta < \frac{3}{2}$	$\frac{1}{2\cos \pi\delta} \frac{\lvert t\rvert^{2\delta-1}}{\Gamma(2\delta)}$
Integrated flicker	2	$\lvert 2\pi f\rvert^{-3}$	$\frac{t^2}{2\pi} \log \lvert t\rvert$
Integrated Wiener	2	$(2\pi f)^{-4}$	$\frac{\lvert t\rvert^3}{12}$

What we are calling the Wiener process is the standard Wiener process plus an arbitrary constant. Similarly, the SI(2) processes have an arbitrary additive constant and slope. We can only compute covariances of linear combinations that are transparent to these ambiguities. For $d = 1$ it may seem strange to generate variances from a negative function, yet (9) works out correctly.

Integrated flicker can be used as a model for the phase of some quartz clocks or hydrogen maser clocks over limited time scales. Here is a simulated sample. To remove the ambiguous constant and slope, a straight line was added to make $X(0) = X(500) = 0$. This process is not mean-square differentiable, and abrupt slope changes occur at all scales. There is a temptation to blame such slope changes on external events.



2 Linear invariant prediction

Let $X \in \mathrm{SI}(d)$ with a known even GACV function $s_X(t)$. Assume that the trend coefficient c_d is 0. Let $T = \{t_1, \ldots, t_p\} \subset \mathbb{R}$, where $p \geq d$, and $t_* \in \mathbb{R}$. We want the linear estimator

$$\hat{X}(t_*) = \sum_{t \in T} a(t) X(t) \tag{10}$$

of $X(t_*)$ that is optimal in the mean-square sense.

If $X(t)$ is stationary with mean zero and ACV $s_X(t)$, we can let $\hat{X}(t_*)$ be the orthogonal projection of $X(t_*)$ on the span of $X(t)$, $t \in T$. Let

$$X(T) = [x(t) : t \in T]^{\mathrm{T}}, \quad a = [a(t) : t \in T]^{\mathrm{T}}$$

(column vectors). Then $\mathrm{E}\left[X(t_*) - a^{\mathrm{T}} X(T)\right] X(T)^{\mathrm{T}} = 0$, giving

$$Va = r, \tag{11}$$

where

$$r = [s_X(t_* - t) : t \in T]^{\mathrm{T}}, \quad V = [s_X(t - u) : t, u \in T].$$

Equation (11), called the Yule-Walker equations, have a solution for a, and any solution (in case V is singular) gives the same $\hat{X}(t_*)$.

Now let $d \geq 1$, so that we can only compute covariances of members of $\mathcal{S}_d$. We adopt two principles that will determine the optimal estimator of form (10).

1 Invariance. The estimation error $X(t_*) - \hat{X}(t_*)$ is to be invariant to any polynomial of degree $\leq d - 1$ that we may add to $X(t)$. In other words,

$$X(t_*) - \hat{X}(t_*) \in \mathcal{S}_d(T \cup \{t_*\}). \tag{12}$$

This statement is equivalent to

$$a^{\mathrm{T}} A = q^{\mathrm{T}}, \tag{13}$$

where

$$A = \begin{bmatrix} 1 & t_1 & \cdots & t_1^{d-1} \\ & & \vdots & \\ 1 & t_p & \cdots & t_p^{d-1} \end{bmatrix}, \quad q = \begin{bmatrix} 1 & t_* & \cdots & t_*^{d-1} \end{bmatrix}^{\mathrm{T}}.$$

Because $p \geq d$, A has rank d. A random variable $a^{\mathrm{T}} X(T)$ with a satisfying (13) is called a *linear invariant estimator* (LIE) of $X(t_*)$. Because $\sum_t a(t) = 1$, any LIE is unbiased. The LIEs form a nonempty affine space $\mathcal{A}_d$, whose set of differences $\mathcal{A}_d - \mathcal{A}_d$ is just $\mathcal{S}_d(T)$. If $p = d$ there is just one LIE, the value at t_* of the interpolating polynomial of $X(t)$, $t \in T$. When $p = d = 2$, linear extrapolation is often used for clock prediction; its mean square error is often close to optimal when it is minimized with respect to $|t_2 - t_1|$.

2 Orthogonality. The optimal estimator $\hat{X}(t_*)$ is to be the LIE that is closest to $X(t_*)$ in L^2. It is not hard to show that a unique best LIE (BLIE) exists and satisfies the orthogonality condition

$$X(t_*) - \hat{X}(t_*) \perp \mathcal{S}_d(T), \tag{14}$$

that is,

$$\mathrm{E}\left[X(t_*) - a^{\mathrm{T}} X(T)\right] \left[X(T)^{\mathrm{T}} b\right] = 0 \tag{15}$$

whenever the entries of $b = [b(t) : t \in T]^{\mathrm{T}}$ kill $\mathcal{P}^{d-1}$. Both factors of (15) being in $\mathcal{S}_d$, we can use the GACV $s_X(t)$ to evaluate it by (3), giving

$$\left(r^{\mathrm{T}} - a^{\mathrm{T}} V\right) b = 0 \text{ whenever } b^{\mathrm{T}} A = 0. \tag{16}$$

Then $r - Va$ must be in the column space of A:

$$r - Va = A\theta \tag{17}$$

for some $\theta = [\theta_0 \ldots \theta_{d-1}]^{\mathrm{T}}$.

We now have the equations

$$A^{\mathrm{T}}a = q, \quad Va + A\theta = r, \tag{18}$$

in $p + d$ unknowns, which generalize the Yule-Walker equations (11). But V is not positive definite; its main diagonal is usually zero. I don't know conditions for V to be nonsingular, but if it is, then (18) has the following solution: let $\Lambda = A^{\mathrm{T}}V^{-1}A$; then

$$\begin{aligned} \theta &= \Lambda^{-1}\left(A^{\mathrm{T}}V^{-1}r - q\right), \\ a &= V^{-1}\left(r - A\theta\right). \end{aligned} \tag{19}$$

Another approach, which I call "brute force", is to write (18) as the single matrix equation

$$\begin{bmatrix} V & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} a \\ \theta \end{bmatrix} = \begin{bmatrix} r \\ q \end{bmatrix}, \tag{20}$$

and tell Matlab to evaluate $M \backslash \left[\begin{smallmatrix} r \\ q \end{smallmatrix}\right]$, where M is the square matrix in (20). Matlab often reports that M is poorly scaled or conditioned (V by itself can be bad enough), yet this method seems to work all right in practice, too.

Having solved for a and θ we can easily evaluate the mean square error from the GACV and (3). Using (18) we have

$$\begin{aligned} \mathrm{E}\left[X(t_*) - \hat{X}(t_*)\right]^2 &= \mathrm{E}\left[X(t_*) - a^{\mathrm{T}}X(T)\right]\left[X(t_*) - X(T)^{\mathrm{T}}a\right] \\ &= s_X(0) - r^{\mathrm{T}}a - a^{\mathrm{T}}r + a^{\mathrm{T}}Va \\ &= s_X(0) - r^{\mathrm{T}}a - q^{\mathrm{T}}\theta. \end{aligned} \tag{21}$$

Initially I assumed $c_d = 0$. If $c_d \neq 0$ but is known, then we can predict the corrected phase $X(t) - c_d t^d/d!$, whose trend is 0. If we don't know c_d, then we are obliged to treat X as a member of $\mathrm{SI}(d+1)$ (if T has at least $d+1$ points). The penalty is a greater mean square error, because we are projecting $X(t_*)$ on a smaller affine space $\mathcal{A}_{d+1}$. For example, if the long-term average phase slope (rate error) of a cæsium clock is known, then the corrected phase is well modeled by a constant times the Wiener process, for which $d = 1$. For the standard Wiener process, the optimal predictor of $X(\tau)$ based on $\{-t, \ldots, 0\}$ $(\tau, t > 0)$ is $X(0)$ with mean square error τ. But if the average slope is uncertain, then we have to use $d = 2$ to get a predictor that is invariant to level and slope. In this case, $\hat{X}(\tau) = \left(1 + \frac{\tau}{t}\right)X(0) - \frac{\tau}{t}X(-t)$ with MSE $\tau\left(1 + \frac{\tau}{t}\right)$. As $t \to \infty$, the two-point predictor recovers the unknown slope, and the MSE tends to τ.

3 Equally spaced data

Solving the Yule-Walker equations (11) or their generalization (18) by general linear equation solving methods requires $O(p^3)$ operations for fixed d. If T is equally spaced, however, then V is a Toeplitz matrix. For stationary $X(t)$, the Levinson-Durbin algorithm [5], which is a loop on p, calculates the regression coefficients and the mean square error in $O(p^2)$ operations. I have been able to extend this algorithm to the $\mathrm{SI}(d)$ case while keeping the $O(p^2)$ property for fixed d, but I will not present this material here. In the cases I have tried, the two general algorithms (V^{-1} and brute force) and the extended Levinson algorithm agree within roundoff error.

Even if the whole data set is equally spaced, the general case is still useful. On examining the regression coefficients $a(t)$ for an equally spaced T, one often sees that most of the $a(t)$ are small. Then, by applying the general method to a subset of T where the $a(t)$ are relatively large, one can verify excellent suboptimal estimators that use only a few of the available data. The ones you use are often bunched near the minimum and maximum of T.

References

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

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

[3] Greenhall C A 1999 The generalized autocovariance: a tool for clock noise statistics *TMO Progress Report* **42-137** Jet Propulsion Laboratory, Pasadena, CA http://tmo.jpl.nasa.gov/tmo/progress_report/42-137/137J.pdf

[4] Mandelbrot B B, Van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Review* **10** 422–437

[5] Brockwell P J, Davis R A 1991 *Time series: theory and methods, second edition* §5.2 Springer

Optimal prediction of clocks from finite data

Charles Greenhall

Jet Propulsion Laboratory, California Institute of Technology

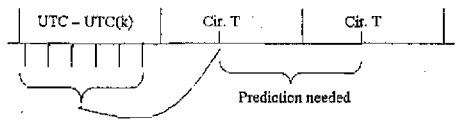
Honoring Adriano Garsia's 75th birthday

- Background: clock prediction, clock noise models
- Linear invariant prediction

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

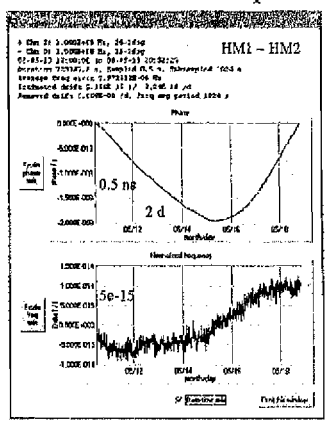
Need for prediction of clock differences

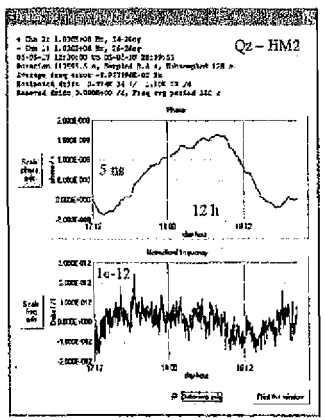
Monthly cycle of BIPM Circular T reports



BIPM = Bureau International des Poids et Mesures, Sèvres
UTC = Coordinated Universal Time
UTC(k) = Laboratory k's version of UTC

Examples of clock noise





General noise model: processes with stationary increments of order d

$d = 1$: Kolmogorov (1940s)
$d \geq 1$: Yaglom, Pinsker (1950s)

A. M. Yaglom, 1958, Correlation theory of processes with random stationary nth increments, AMS Translations, Series 2, Volume 8, pp. 87-141.

Stationary processes (real-valued)

$X(t), t \in \mathbb{R}$, continuous in $L^2(\Omega)$,

$\mathrm{E}X(t)X(u) = s_X(t-u)$; $s_X(t)$ = autocovariance (ACV) function

$$s_X(t) = \mathrm{E}C^2 + \int_{-\infty}^{\infty} e^{i2\pi ft} S_X(f)\,df$$

$S_X(f)df$ = spectral measure, finite, no atom at $f = 0$

$$C = \lim_{T\to\infty} \frac{1}{T}\int_0^T X(t)\,dt \text{ in } L^2 = \text{ mean (possibly random)}$$

$a(t)$: "finite window", nonzero on a finite set, $\tilde{a}(f) = \sum_t a(t)e^{i2\pi ft}$

$$\mathrm{E}\left[\sum_t a(t)X(t)\right]\left[\sum_t b(t)X(t)\right] = \sum_{t,u} a(t)b(u)s_X(t-u)$$

$$= \mathrm{E}C^2\left[\sum a(t)\right]\left[\sum b(t)\right] + \int_{-\infty}^{\infty} \tilde{a}(f)\overline{\tilde{b}(f)}S_X(f)\,df$$

Processes with stationary dth increments

$X \in \mathrm{SI}(d)$: For every τ, the process $\Delta_\tau^d X(t) = \sum_{k=1}^{d}\binom{d}{k}(-1)^k X(t-k\tau)$ is stationary.

Let $C_d = \lim_{T\to\infty}\frac{1}{T}\int_0^T \Delta_t^d X(t)\,dt$ in L^2 : trend coefficient (0 for $X(t) - \frac{C_d t^d}{d!}$)

$\mathrm{SI}(0) \subset \mathrm{SI}(d) \subset \mathrm{SI}(d+1)$, $X \in \mathrm{SI}(d) \Rightarrow C_{d+1} = 0$, $\int dt : \mathrm{SI}(d) \to \mathrm{SI}(d+1)$.

Yaglom derived a spectral measure $S_X(f)df$ on $\mathbb{R} - \{0\}$, even, such that

$$\int_0^1 f^{2d} S_X(f)\,df < \infty, \quad \int_1^\infty S_X(f)\,df < \infty.$$

Say that $a(t)$ satisfies the moment condition for d if $\sum_t a(t)t^j = 0$, $0 \le j \le d-1$.

Then

$$\mathrm{E}\left[\sum_t a(t)X(t)\right]\left[\sum_t b(t)X(t)\right]$$

$$= \mathrm{E}C_d^2\left[\sum a(t)\frac{t^d}{d!}\right]\left[\sum b(t)\frac{t^d}{d!}\right] + \int_{-\infty}^{\infty} \tilde{a}(f)\overline{\tilde{b}(f)}S_X(f)\,df$$

whenever $a(t), b(t)$ satisfy the moment condition for d.

Generalized autocovariance (GACV)

Define $s_X(t) = \mathrm{E}C_d^2\frac{(-t^2)^d}{(2d)!} + \int_{|f|\le 1}\left[e^{i2\pi ft} - \sum_{k=0}^{2d-1}\frac{(i2\pi ft)^k}{k!}\right]S_X(f)\,df$

$$+ \int_{|f|>1} e^{i2\pi ft} S_X(f)\,df \quad \text{mod } P^{2d-1}.$$

Then $s_X(t)$ is continuous; may assume $s_X(-t) = s_X(t)$.

If $a(t), b(t)$ satisfy the moment condition for d, then

$$\mathrm{E}\left[\sum_t a(t)X(t)\right]\left[\sum_t b(t)X(t)\right] = \sum_{t,u} a(t)b(u)s_X(t-u)$$

Example: "structure function" for $d = 1$

$$D(\tau) := \mathrm{E}\left[X(t+\tau) - X(t)\right]^2 = 2s_X(0) - s_X(\tau) - s_X(-\tau) = \left(-\delta_\tau^2 s_X\right)(0)$$

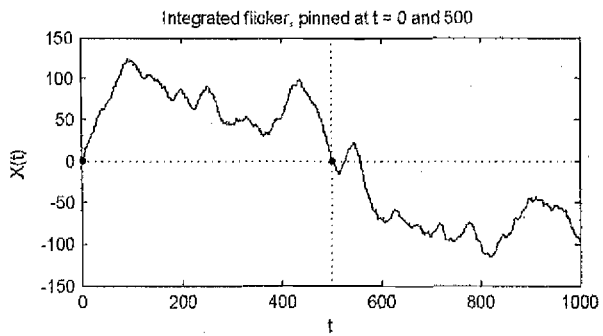
More generally, ACV of stationary process $\Delta_\tau^d X(t)$:

$$\mathrm{E}\Delta_\tau^d X(t)\Delta_\tau^d X(u) = (-1)^d\left(\delta_\tau^{2d} s_X\right)(t-u)$$

Examples of spectrum and GACV

Name	d	$S_X(f)$	$s_X(t)$
Wiener	1	$(2\pi f)^{-2}$	$-\frac{\lvert t\rvert}{2}$
Fractional BM	1	$\lvert 2\pi f\rvert^{-2\delta}$, $\frac{1}{2} < \delta < \frac{3}{2}$	$\frac{\lvert t\rvert^{2\delta-1}}{2\cos\pi\delta\,\Gamma(2\delta)}$
Integrated flicker	2	$\lvert 2\pi f\rvert^{-3}$	$\frac{t^2\log\lvert t\rvert}{2\pi}$
Integrated Wiener	2	$\lvert 2\pi f\rvert^{-4}$	$\frac{\lvert t\rvert^3}{12}$

Sample of simulated integrated flicker (f^{-3} noise)



Linear invariant prediction

Given $X \in \mathrm{SI}(d)$, $s_X(t)$ known, $C_d = 0$.

Given $X(t)$ on $T = \{t_1, \ldots, t_p\}$, $p \geq d$, estimate $X(t_*)$ for some other t_*.

Want linear estimator $\hat{X}(t_*) = \sum_{t \in T} a(t) X(t)$, optimal in L^2 sense.

Stationary case, $d = 0$

Project $X(t_*)$ orthogonally on $X(t)$, $t \in T$. Let $X(T) = \begin{bmatrix} X(t_1) \\ \vdots \\ X(t_p) \end{bmatrix}$, $a = \begin{bmatrix} a(t_1) \\ \vdots \\ a(t_p) \end{bmatrix}$

Then $\mathrm{E}\left[X(t_*) - a^T X(T)\right] X(T)^T = 0$, i.e.,

$$Va = r$$

(Yule-Walker equations), where

$$V = \mathrm{E}X(T)X(T)^T = \left[s_X(t_i - t_j)\right], \quad r = \mathrm{E}X(t_*)X(T) = \begin{bmatrix} s_X(t_* - t_1) \\ \vdots \\ s_X(t_* - t_p) \end{bmatrix}$$

Here, V is a covariance matrix.

The spaces $\mathcal{S}_d$

$\mathcal{S}_d$: The (non-closed) space of random variables $\sum_t b(t) X(t)$

where $b(t)$ is a finite window satisfying the moment condition for d.

$\mathcal{S}_d(T)$: Same with $b(t)$ supported on the finite set T.

Can use GACV $s_X(t)$ to calculate inner product of any two elements of $\mathcal{S}_d$:

$$\mathrm{E}\left[\sum_t b_1(t) X(t)\right]\left[\sum_t b_2(t) X(t)\right] = \sum_{t,u} b_1(t) b_2(u) s_X(t-u)$$

where $b_1(t), b_2(t)$ satisfy the moment condition for d.

Invariance principle for predictor ($d \geq 1$)

Conditions for $\hat{X}(t_*) = a^T X(T)$ as estimate of $X(t_*)$

1. Invariance of error to polynomials of degree $\leq d-1$:

$$X(t_*) - \hat{X}(t_*) \in \mathcal{S}_d\left(T \cup \{t_*\}\right), \text{ i.e.,}$$

$$\sum_{t \in T} a(t) t^j = t_*^j, \quad j = 0, \ldots, d-1, \text{ i.e.,}$$

$$a^T A = q^T, \text{ where}$$

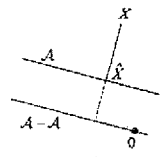
$$A = \begin{bmatrix} 1 & t_1 & \cdots & t_1^{d-1} \\ & & \vdots & \\ 1 & t_p & \cdots & t_p^{d-1} \end{bmatrix}, \quad q^T = \begin{bmatrix} 1 & t_* & \cdots & t_*^{d-1} \end{bmatrix}$$

Such an $\hat{X}(t_*)$ called a linear invariant estimator (LIE) of $X(t_*)$.

The LIEs form an affine subset $\mathcal{A}$ of L^2, and $\mathcal{A} - \mathcal{A} = \mathcal{S}_d(T)$.

For $p = d$, $\exists!$ LIE: extrapolation of interpolating polynomial.

Orthogonality principle



2. Orthogonality of error to $\mathcal{A} - \mathcal{A}$:

$$X(t_*) - \hat{X}(t_*) \perp \mathcal{S}_d(T), \text{ i.e.,}$$

$$\mathrm{E}\underbrace{\left[X(t_*) - a^{\mathrm{T}} X(T)\right]}_{\in \mathcal{S}_d}\underbrace{\left[X(T)^{\mathrm{T}} b\right]}_{\in \mathcal{S}_d} = 0$$

whenever b satisfies the moment condition for d.

Can evaluate with GACV $s_X(t)$:

$$\left(r^{\mathrm{T}} - a^{\mathrm{T}} V\right) b = 0 \text{ whenever } b^{\mathrm{T}} A = 0.$$

Then for some d-vector θ,

$$r - Va = A\theta$$

Result: $p + d$ equations in $p + d$ unknowns a, θ:

$$A^{\mathrm{T}} a = q, \quad Va + A\theta = r$$

Here, $V = \left[s_X(t_i - t_j)\right]$ is not a covariance matrix.

Solution methods for BLIE

(Best linear invariant estimator)

$$A^{\mathrm{T}} a = q, \quad Va + A\theta = r$$

If V is nonsingular, let $\Lambda = A^{\mathrm{T}} V^{-1} A$; then

$$\theta = \Lambda^{-1}\left(A^{\mathrm{T}} V^{-1} r - q\right)$$

$$a = V^{-1}\left(r - A\theta\right)$$

Brute-force approach: write system as

$$\begin{bmatrix} V & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} a \\ \theta \end{bmatrix} = \begin{bmatrix} r \\ q \end{bmatrix}$$

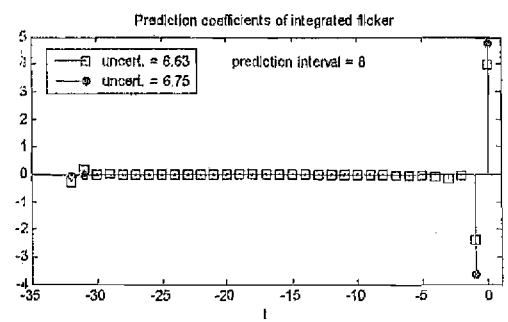
and solve with one Matlab statement (complains).

Mean square error of BLIE

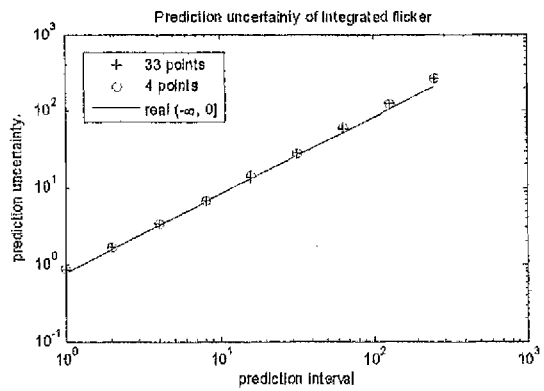
$$\begin{aligned}
\mathrm{E}\left[X(t_*) - \hat{X}(t_*)\right]^2 &= \mathrm{E}\left[X(t_*) - a^{\mathrm{T}} X(T)\right]\left[X(t_*) - X(T)^{\mathrm{T}} a\right] \\
&= s_X(0) - r^{\mathrm{T}} a - a^{\mathrm{T}} r + a^{\mathrm{T}} V a \\
&= s_X(0) - r^{\mathrm{T}} a - q^{\mathrm{T}} \theta
\end{aligned}$$

For $d \geq 1$, $s_X(0)$ is often 0, and the separate terms are meaningless as covariances.

Integrated flicker, prediction coefficients



Integrated flicker, prediction uncertainty



Equally spaced data

$$\text{Predictor: } \hat{X}_{n+m} = \sum_{k=1}^{p} \phi_k(p) X_{n-k}, \quad m \geq 0$$

Stationary case.

Matrix V of Yule-Walker equations is Toeplitz.

Levinson-Durbin algorithm iterates $p-1 \to p$ in $O(p)$ steps,

giving $O(p^2)$ solution instead of $O(p^3)$.

SI(d) case.

Found extension of Levinson algorithm that preserves $O(p^2)$ property,

with some overhead depending on d. Uses θ from previous batch equations.