Correspondence_____

A Note on the Prediction Error for Small Time Lags into the Future

JAMES A. BUCKLEW, MEMBER, IEEE

Abstract—Explicit expressions are derived for derivatives at zero lag of the mean-square prediction error for a class of random processes that includes those with rational power spectra. A simple random process not in the class is demonstrated. A sufficient condition for membership in the class is given.

I. INTRODUCTION

Finding the minimum variance linear predictor of a continuous-time, wide-sense-stationary random process is a classical problem in the electrical engineering and mathematics literature. Wiener [1] solved the problem of prediction given that the infinite past of the process is observed. Krein [2] solved the problem in principle for the much harder case of observing only a finite portion of the entire past. A difficulty is that the prediction formula is difficult to evaluate in practice. Rozanov [3] gives formulas for the rational power-spectrum case, but these still require solving complex differential equations to obtain certain constants. Alternatively, one can use the complex analysis approach of Yaglom [4], which is similar in computational difficulty. Given that the predictors may be difficult to calculate, one may choose to resort to suboptimal ad hoc schemes. If so it is of crucial importance to be able to calculate the mean-square error performance of such a suboptimal predictor and compare it to the performance of the optimal one.

Cuzick [5] gives a nice overview of the problem and derives an upper bound to the prediction error that approaches zero at the correct rate as the "lag" or time interval into the future approaches zero. This bound depends on an unknown constant that hampers (as the author states) the practical utility of the results.

An idea to be investigated here, for the small time-lag prediction of a continuous random process, is the use of Taylor series to extrapolate into the future from the last point observed. If only m mean-square derivatives of the random process are available, then the series is truncated after those terms. One would imagine that as the time lag goes to zero, this should be a very good estimator of the future. This reasoning turns out to be essentially correct for a large class of random processes. More interesting (perhaps unfortunately), this is an incorrect way of thinking about some random processes. In other words, Taylor series predictors can be very bad in comparison to optimal prediction even as the time lag approaches zero.

We will give approximate expressions for the mean-squared prediction error for small extrapolations for a class of random processes that includes those with rational power spectra. Much of our intuition is based upon how filtering and prediction works for the rational case. As a further interesting sidelight, it is demonstrated that the predictors (and their error) can depart very widely from those that would be expected based upon a rational approximation.

Notation

$\{x(t) \quad -\infty \le -T \le t \le 0\}$	An observation of a wide-sense-stationary nondeterministic zero-mean random process.
$x^{(k)}(t)$	The kth mean-square derivative of $x(t)$ (if it exists).
$\hat{x}_T(\tau)$	The minimum mean-squared error predictor of $x(\tau)$, $\tau \ge 0$.
$E\{(x(\tau) - \hat{x}_T(\tau))^2\}$	$\triangleq mse_T(\tau)$.
$F'(\lambda)$	The power spectrum of the random process $x(t)$; $\mathscr{F}^{-1}\{F'(\lambda)\}$ $= R(t)$, the autocorrelation function the random process where $\mathscr{F}\{\cdot\}$ and $\mathscr{F}^{-1}(\cdot)$ denote the Fourier and inverse Fourier Transform operators, respectively.
$R(t)$	$= f(t) \cdot f(-t)$, where $f(t)$ is a causal square integrable function that is guaranteed to exist by the Paley–Wiener theorem [6].

II. DEVELOPMENT

Suppose $F'(\lambda)$ is rational and $F'(\lambda) \ne 0$; then $\mathscr{F}\{f(t)\}$ is also rational. Then we have

$$F'(\lambda) = \left| \frac{\displaystyle\sum_{j=0}^{\alpha} A'_j(i\lambda)^j}{\displaystyle\sum_{j=0}^{\beta} B'_j(i\lambda)^j} \right|^2 ,$$

where A'_j and B'_j are real, and $\beta \ge \alpha + 1$, and

$$\mathscr{F}\{f(t)\} = \frac{\displaystyle\sum_{j=0}^{\alpha} A'_j(i\lambda)^j}{\displaystyle\sum_{j=0}^{\beta} B'_j(i\lambda)^j} .$$

Let k be such that $\beta \ge \alpha + k + 1$. Then the Laplace transform of the kth derivative of $f(t)$ is given by

$$\mathscr{L}\{f^{(k)}(t)\} = \frac{(s^k) \displaystyle\sum_{j=0}^{\alpha} A'_j\left(\frac{s}{2\pi}\right)^j}{\displaystyle\sum_{j=0}^{\beta} B'_j\left(\frac{s}{2\pi}\right)^j} .$$

By the Abelian theorem for Laplace transforms we have

$$f^{(k)}(0) = \lim_{s \to \infty} \mathscr{L}\{f^{(k)}(t)\} s$$

$$= 0, \qquad \beta > \alpha + k + 1$$

$$= \frac{A'_\alpha}{B'_\beta}(2\pi)^{k+1}, \qquad \beta = \alpha + k + 1.$$

However, from [7, p. 544] we have

$$R^{(k)}(0^+) = 0, \qquad 2\beta \ge 2\alpha + k + 2$$

$$= +\frac{1}{2}\left|\frac{A'_\alpha}{B'_\beta}\right|^2 (2\pi i)^{2\beta - 2\alpha}, \qquad 2\beta = 2\alpha + k + 1.$$

Hence

$$f^{(k)}(0)^2 = 2R^{(2k+1)}(0^+)(-1)^{k+1}, \quad \text{for } \beta - \alpha - 1 \geq k.$$

(1)

A random process of the above type has exactly $\beta - \alpha - 1$ derivatives. Therefore we can consider a suboptimal predictor $\bar{x}(\tau)$, of $x(\tau)$,

$$\bar{x}(\tau) \triangleq \sum_{k=0}^{\beta - \alpha - 1} \frac{\tau^k}{k!} x^{(k)}(0).$$

We define

$$\overline{mse}(\tau) \triangleq E\left\{ (x(\tau) - \bar{x}(\tau))^2 \right\}.$$

It is straightforward (but tedious) to verify that

$$\frac{\partial^{(k)} \overline{mse}(\tau)}{\partial \tau^k} \bigg|_{\tau=0^+} = 2 \frac{\binom{k-1}{\frac{k-1}{2}}}{2} R^{(k)}(0^+)(-1)^{(k+1)/2},$$

$$k \leq 2(\beta - \alpha) - 1.$$

A well-known formula [6], [7] for the optimum prediction when $T = \infty$ is

$$mse_\infty(\tau) = E\left\{ (x(\tau) - \hat{x}_\infty(\tau))^2 \right\} = \int_0^\tau f^2(t) \, dt.$$

Again it is straightforward but tedious to verify that

$$\frac{\partial^k mse_\infty(\tau)}{\partial \tau^k} \bigg|_{\tau=0^+} = \frac{\binom{k-1}{\frac{k-1}{2}}}{2} \left(f^{((k-1)/2)}(0^+) \right)^2,$$

$$k \leq 2(\beta - \alpha) - 1$$

$$= \frac{\partial^{(k)} \overline{mse}(\tau)}{\partial \tau^k} \bigg|_{\tau=0^+}.$$

(We emphasize again that the above is zero unless $k = 2(\beta - \alpha) - 1$). Hence, since $\overline{mse}(\tau)$ represents the error of a suboptimal estimator that depends only upon the value of the random process and its derivatives at zero, we must have

$$k! \frac{\overline{mse}(\tau)}{\tau^k} \geq k! \frac{mse_T(\tau)}{\tau^k} \geq k! \frac{mse_\infty(\tau)}{\tau^k},$$

where $k \leq 2(\beta - \alpha) - 1$. Taking the limit as $\tau \to 0^+$ we have the following theorem.

Theorem 1: Suppose the random process $x(t)$ has a rational power spectrum $F'(\lambda)$ such that $F'(\lambda) \neq 0$,

$$\frac{\partial^k mse_T(\tau)}{\partial \tau^k} \bigg|_{\tau=0^+} = 2 \frac{\binom{k-1}{\frac{k-1}{2}}}{2} R^{(k)}(0^+)(-1)^{(k+1)/2},$$

$$k \leq 2(\beta - \alpha) - 1,$$

where $R(\tau) = \mathscr{F}^{-1}\{F'(\lambda)\}$.

Remark 1: The above development proves the existence of the partial derivative and gives its value at $\tau = 0^+$. For all values of k smaller than $2(\beta - \alpha) - 1$ the value is zero.

Remark 2: A wide-sense-stationary random process has m mean-square continuous derivatives if and only if its autocorrelation function possesses $2m$ continuous derivatives at zero. From a direct construction using the Taylor series error expression one can show that if $R(\tau)$ is continuously differentiable m times at the origin, then $\partial^k mse_T(\tau)/\partial \tau^k|(\tau = 0^+) = 0$ for $k = 0, 1, \cdots, m$. It is the $k = m + 1$ case that is of most concern.

Remark 3: By the nature of the proof we see that for small prediction lags the simple Taylor series predictor performs nearly as well as the more complex Wiener predictor for processes with rational power spectra.

We now look for a more general class of random processes that behave in the same manner. Since the rationals are dense in the class of all spectra we might hope that all random processes would behave like this. This is not the case, as shown by the following counterexample.

Counterexample: Suppose $F'(\lambda) = \sin^2(\lambda/2)/\lambda^2$ and $T/2 = N$ a positive integer. Then from [6] we have

$$mse_T(\tau) = \frac{2N+2}{8(2N+1)} \tau + \frac{(N+1-\tau)\tau}{4(2N+1)},$$

$$0 \leq \tau < 1/2$$

$$\frac{\partial mse_T(\tau)}{\partial \tau} \bigg|_{\tau=0^+} = \frac{2N+2}{(2N+1)8} + \frac{N+1}{(2N+1)4}$$

$$= \frac{N+1}{(2N+1)2}.$$

Suppose we use $\bar{x}(\tau) = R(\tau)/R(0)x(0)$ (the best predictor based upon $x(0)$). Then since

$$R(\tau) = \begin{cases} \dfrac{1}{4} - \dfrac{1}{4}\tau, & \text{for } 0 \leq \tau \leq 1, \\ 0, & \text{for } \tau > 1 \end{cases}$$

$$\overline{mse}(\tau) = \frac{1}{4} - \frac{\left(\dfrac{1}{4} - \dfrac{1}{4}\tau \right)^2}{1/4}$$

$$\frac{\partial \overline{mse}(\tau)}{\partial \tau} \bigg|_{\tau=0^+} = \frac{1}{2}.$$

Remark 4: Obviously, these two expressions are unequal, with the optimal predictor error derivative strictly less than $1/2$ for all integer values of N and approaching $1/4$ as N approaches ∞.

Remark 5: In contrast to the rational case, the partial derivative of the error depends upon the length of the observation T we are given.

If we impose some technical conditions, we can demonstrate an entire class of power spectra that share the desired property with the rational case.

Theorem 2: Suppose $f(t)$ has the representation

$$f(t) = f(0) + \int_0^t f'(\alpha) \, d\alpha,$$

where $f'(\alpha)$ is square integrable, i.e., $f(t)$ is absolutely continuous and its derivative is a square integrable function. Then $R'(0^+) = -f^2(0)/2$.

Proof: For $\tau > 0$ we have

$$R(\tau) = \int_\tau^\infty f(\alpha) f(\alpha - \tau) \, d\alpha$$

$$= \int_\tau^\infty f(\alpha) \left[f(\alpha) - \int_0^\tau f'(\alpha - s) \, ds \right] d\alpha,$$

by absolute continuity of f

$$= \int_\tau^\infty f^2(\alpha) \, d\alpha - \int_0^\tau \int_0^\infty f(\alpha) f'(\alpha - s) \, d\alpha \, ds$$

$$+ \int_0^\tau \int_0^\tau f(\alpha) f'(\alpha - s) \, d\alpha \, ds.$$

(The integration interchange is justified by invoking Tonelli's theorem on $|f(\alpha)f'(\alpha - s)|$ to get integrability on cross measure;

then Fubini's theorem allows interchange.) Now consider the last term divided by τ, i.e.,

$$\frac{1}{\tau}\int_0^\tau \int_0^\tau f(\alpha)f'(\alpha - s) \, d\alpha \, ds$$

$$\leq \frac{1}{\tau}\int_0^\tau \left(\int_0^\tau f^2(\alpha) \, d\alpha\right)^{1/2} \left(\int_0^\tau f'^2(\alpha - s) \, d\alpha\right)^{1/2} ds,$$

by Schwartz inequality

$$\leq \frac{1}{\tau}\int_0^\tau K\left(\int_0^\tau f^2(\alpha) \, d\alpha\right)^{1/2} ds$$

$$= \frac{1}{\tau}\tau K (mse_\infty(\tau))^{1/2} = K mse_\infty(\tau)^{1/2} \underset{\tau\to 0^+}{\to} 0.$$

Therefore we may invoke the fundamental theorem of the calculus and find (for $\tau > 0$)

$$R'(\tau) = -f^2(\tau) - \int_0^\infty f(\alpha)f'(\alpha - \tau) \, d\alpha + 0(\tau),$$

where we have terms that are approaching zero as τ approaches zero by the above argument. Utilizing the fact that f' is square integrable, we may uniformly approximate it by a continuous function (alternatively use Parseval's theorem, since $f, f' \in L^2$). This implies that we can take the limit as τ goes to zero through positive values to obtain

$$R'(0^+) = -f^2(0) - \int_0^\infty f(\alpha)f'(\alpha) \, d\alpha$$

$$= -f^2(0) - \left.\frac{f^2(\alpha)}{2}\right|_0^\infty$$

$$= -f^2(0) + \frac{f^2(0)}{2} = -\frac{f^2(0)}{2}.$$

The next to last line follows since

$$\int_0^\infty |f(x)||f'(x)| \, dx = \lim_{T\to\infty}\int_0^T |f(x)||f'(x)| \, dx$$

$$\leq \|f\|\|f'\| = M < \infty.$$

This implies the existence of the limit

$$\lim_{T\to\infty}\int_0^T f(x)f'(x) \, dx = M' < \infty,$$

because absolute summability implies summability. Therefore $\lim_{T\to\infty} f(T)^2$ exists and hence must equal zero.

Remark 6: The theorem statement corresponds to (1) with $k = 0$. We could obtain a statement for arbitrary k by making the same assumptions on $f^{(k-1)}(t)$ that we do for $f(t)$. Although the proof is messier, it is essentially the same.

Remark 7: The counterexample fails the theorem requirements since

$$f(t) = \frac{1}{2}, \qquad 0 \leq t \leq 1.$$

We obtain $f'(t) = (-1/2)\delta(t - 1)$, and this (generalized) derivative is not square integrable.

Remark 8: One can generate entire families of power spectra that behave as the rational spectra do. For example, take $f(t) = 0$ for $t < 0$ and $f(t) = \min\{1, e - (t - 1)\}$ for $t \geq 0$. Then $f(t) = 1 + \int_0^t f'(\alpha) \, d\alpha$. Then $f'(t) = -e^{-(t-1)}$ for $t > 1$ and $f'(t) = 0$ for $t \leq 1$, which is square integrable. Then

$$R_x(\tau) = 2 - \tau - \frac{e^{-\tau}}{2}, \qquad 0 \leq \tau \leq 1,$$

$$\left.\frac{\partial R_x(\tau)}{\partial \tau}\right|_{\tau=0} = -\frac{1}{2} = -\frac{f(0)^2}{2}.$$

Remark 9: If $f \in L^2$ and $f' \in L^2$, a well-known property of convolutions states that $f \cdot f'$ is not only continuous but uniformly continuous. This implies that the derivative of the autocorrelation function must exist at every point (see the proof of Theorem 2) and be continuous (except possibly at the origin). This condition would usually be easier to check on a particular autocorrelation function than the condition $f, f' \in L^2$. Our counterexample with the triangular autocorrelation fails this test.

III. CONCLUSIONS

This correspondence concerns a property of the prediction error for small lags. For the rational power spectrum case a closed-form expression is derived for the derivative with respect to the lag of the mean-square prediction error. It is shown that a large class of "nice" power spectra behaves like the rationals, but that there are also simple ones, in particular the triangular autocorrelation function that do not. For these nice random processes, simple Taylor series-type predictors perform optimally well for small lags into the future.

Whether a particular power spectrum is nice or not remains an open problem. A sufficient condition, in terms of smoothness attributes of the functions in the Wiener spectral decomposition has been given, but its utility is limited due to the difficulty of performing the decomposition and verifying the conditions. We can say that if the autocorrelation function is not everywhere differentiable (except possibly at zero), then it will not satisfy our smoothness conditions.

REFERENCES

[1] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Times Series.* New York: Wiley, 1949.
[2] M. Krein, "On a fundamental approximation problem in the theory of extrapolation and filtering of stationary random processes," *Dokl. Adad. Nauk. SSSR,* vol. 94, pp. 13–16, 1954 (English translation in *Select Transl. Math. Statist. Probab.,* vol. 4, pp. 127–131, 1964.
[3] Y. Rozanov, *Stationary Random Processes.* San Francisco: Holden-Day, 1967.
[4] A. Yaglom, "Effective solutions of linear approximation problems for multi-variate stationary processes with a rational spectrum," *Theory Probab. Appl.,* pp. 239–264, vol. 5, 1960.
[5] J. Cuzick, "A lower bound for the prediction error of stationary Gaussian processes," *Indiana Univ. Math. J.,* vol. 26, no. 3, 1977.
[6] H. Dym and H. McKean, *Gaussian Processes, Function Theory, and the Inverse Spectral Problem.* New York: Academic, 1976.
[7] J. Doob, *Stochastic Processes.* New York: Wiley, 1953.
[8] H. Royden, *Real Analysis.* Toronto: Macmillan, 1968.