International Journal of Systems Signal Control and Engineering Application

MEDWELL PUBLICATIONS SINCE 2000

Non-Recursive Prediction of Random Processes

Vladimir Alekseevich Golovkov

Institute of Optoelectronic Instrumentation Engineering, Leningrad Region, 188540 Sosnovy Bor, Russia

Key words: stochastic process, prediction, correlation function, prediction evaluation variance, sample efficiency

Corresponding Author:
Vladimir Alekseevich Golovkov
Institute of Optoelectronic Instrumentation Engineering, Leningrad Region, 188540 Sosnovy Bor, Russia

Page No.: 117-120
Volume: 10, Issue 4, 2017
ISSN: 1997-5422
International Journal of Systems Signal Control and Engineering Application

Abstract: The study suggests analytical expressions for algorithms of optimal linear prediction of a random process relyingon a sample of the process values and values of its derivatives at a previous instant of time. I also investigate relative efficiency of such algorithms in comparison to transversal algorithms, exemplified by a stochastic process with a finite correlation function.

INTRODUCTION

Let us consider the algorithm and efficiency of linear prediction of a stationary differentiable random process $\xi(t)$ with a correlation function $R(\tau) = \sigma^2\rho(\tau)$ where σ^2 is random process variance and $\rho(\tau)$-a normalized correlation function. Wiener-Hopf transversal filtering (Widrow and Stearns, 1985) implies the use of sample $||X|| = [\xi(t), \xi(t-\Delta), ..., \xi(t-n\Delta t)]^T$ where T-transposition operation, Δt-a time interval between random process counts, n-integer. Sample $||X||$ is assumed to be used for process $\xi(t+\theta)$ forecast evaluation where θ-prediction look-ahead time. On the other hand, for the forecasting we are allowed to use both samples $||X||$ and $||Y|| = [\xi(t), \xi'(t), ..., \xi^{(n)}(t)]^T$ or a sample of the random process and n of its first derivatives which implies random process different ability. Hard warily such sampling is attained via differentiating circuits (Faulkenberry, 1982) or by the well-known finite difference procedure. In this study we will build forecasting algorithms and compare the efficiency of samples ||x||и||Y|| in linear Wiener-Hopf filtering.

MATERIALS AND METHODS

The most accessible method of prediction is linear forecasting or assessment of futurities using linear regression like:

$$\xi(t+\theta/t) = k_0\xi(t)+k_1\xi(t-\Delta t)+, ..., k_n\xi(t-n\Delta t) = \|K\|\cdot\|X\| \quad (1)$$

where, $\xi(t+\theta/t)$-prediction of realization $\xi(t)$ for the time instant $t+\theta$ from time instant t, vector, $||K|| = ||k_0, k_1, ..., k_n||$ components of vector $||K||$ depending on forecast look-ahead time θ, on correlation properties of a random process and delay time interval Δt. Similarly when using the values of variables:

$$\xi(t+\theta/t) = w_{0n}\xi(t)+w_{1n}\xi'(t)+, ..., w_{nm}\xi^{(n)}(t) = \|W\|\cdot\|Y\|^T \quad (2)$$

where, vector $||W|| = ||w_{0,\ n}, w_{1,\ n}, ..., w_{n,\ n}||$ Weighting factors of vector $||W||$ are defined as: w_{ij} where i-order of realization derivative, j-size of the sample used for prediction. Optimal factors w_{ij} for vector ||W|| are dependent on forecast look-ahead time θ and correlation properties of the random process. Both types of filtering are filters having impulse response that is finite temporally. In contrast to recursive filters, non-recursive ones are essentially stable (Widrow and Stearns, 1985). According to Widrow and Stearns (1985), the optimal wiener vector of weighting factors is defined as follows:

$$|G_{opt}\| = \| R \|^{-1} \times \| P \|^{T} \quad (3)$$

where, ||R|| is a cross variance matrix of the elements of a sample of a random proces's previous values ||X|| = ξ(t), ξ(t-Δt), ..., ξ(t-nΔt)||, from which it follows that $\|G_{opt}\|$ = $\|K_{opt}\|$ or a cross-variance matrix of a sample of the values and derivatives of the random process ||Y|| = ||ξ(t), ξ'(t), ..., $\xi^{(n)}(t)$||, then $\|G_{opt}\| = \|W_{opt}\|$. Vector ||P|| is a row-vector of cross-variance of the predicted value of the realization of random process ξ(t+θ/t) and sample elements ||X|| or ||Y||. Minimal forecast evaluation variance ξ(t+θ/t) for samples ||X|| or ||Y|| can be defined, accordingly as:

$$\sigma^2\left[\xi(t+\theta/t)/Y(t)\right] = \sigma^2 - \|P\|\cdot\|R\|^{-1}\|P\|^T = \sigma^2 - \|P\|\cdot\|K_{opt}\| \quad (4)$$

or:

$$\sigma^2\left[\xi(t+\theta/t)/X(t)\right] = \sigma^2 - \|P\|\cdot\|R\|^{-1}\|P\|^T = \sigma^2 - \|P\|\cdot\|W_{opt}\| \quad (5)$$

where, σ^2-variance of process ξ(t). We can determine coefficients $\|K\| = \|k_0(\theta, \Delta t), k_1(\theta, \Delta t), ..., k_n(\theta, \Delta t)\|$ for a transversal filter in the form of $\|K_{opt}\| = \|R\|^{-1}\|P\|^T$ where:

$$\|R\| = \sigma^2 \begin{Vmatrix} 1 & \rho(\Delta t) & \cdots & \rho(n\Delta t) \\ \rho(\Delta t) & 1 & \cdots & \rho\left[(n-1)\Delta t\right] \\ \cdots & \cdots & \cdots & \cdots \\ \rho(n\Delta t) & \rho\left[(n-1)\Delta t\right] & \cdots & 1 \end{Vmatrix} \quad (6)$$

$$\|P\| = \sigma^2 \left\|\rho(\theta),\ \rho(\theta+\Delta t), ..., \rho(\theta+n\Delta t)\right\| \quad (7)$$

Matrix ||R|| and vector ||P|| when using sample ||Y||, can be constructed relying on random process overshoot theory (Tikhonov and Khimenko, 1987). Leaving out intermediary mathematical developments, we obtain ||R|| and ||P|| used for vector Y(t) forecasting up to the fourth derivative inclusive as follows:

$$\|R\| = \sigma^2 \begin{Vmatrix} 1 & 0 & \rho''(0) & 0 & \rho^{(4)}(0) \\ 0 & -\rho''(0) & 0 & -\rho^{(4)}(0) & 0 \\ \rho''(0) & 0 & \rho^{(4)}(0) & 0 & \rho^{(6)}(0) \\ 0 & -\rho^{(4)}(0) & 0 & -\rho^{(6)}(0) & 0 \\ \rho^{(4)}(0) & 0 & \rho^{(6)}(0) & 0 & \rho^{(8)}(0) \end{Vmatrix} \quad (8)$$

$$\|P\| = \sigma^2 \left\|\rho(\theta),\ -\rho'(\theta),\ \rho''(\theta),\ -\rho^{(3)}(\theta),\ \rho^{(4)}(\theta)\right\| \quad (9)$$

As can be seen from Eq. 8, the correlation matrix is sparse with many zero-elements. Using Eq. 3 we can obtain prediction algorithms from sample ||Y|| in an explicit form, up to the two highest derivatives inclusive:

$$\xi(t+\theta/t) = \rho(\theta)\cdot\xi(t) \quad (10)$$

$$\xi(t+\theta/t) = \rho(\theta)\xi(t) - \frac{\rho'(\theta)}{\left[-\rho''(0)\right]}\xi'(t) \quad (11)$$

$$\xi(t+\theta/t) = \frac{\rho(\theta)\rho^{(4)}(0) - p''(\theta)\rho''(0)}{\rho^{4}(0) - \rho''(0)^2}\xi(t) - \frac{\rho'(\theta)}{\left[-\rho''(0)\right]}\xi'(t) + \frac{\rho''(\theta) - \rho(\theta)\rho''(0)}{\rho^{(4)}(0) - \rho''(0)^2}\xi''(t) \quad (12)$$

Estimate variance ξ(t+θ/t) using Eq. 5 can be obtained if we use the sample of random process values up to its two first derivatives in the following form:

$$\sigma^2\left[\xi(t+\theta)/\xi(t)\right] = \sigma^2\left\lfloor 1-\rho(\theta)^2 \right\rfloor \quad (13)$$

$$\sigma^2\left[\xi(t+\theta)/\xi(t),\ \xi'(t)\right] = \left[1-\rho(\theta)^2 - \frac{\rho'(\theta)^2}{\left[-\rho''(0)\right]}\right] \quad (14)$$

$$\sigma^2\left[\xi(t+\theta)/\xi(t),\ \xi'(t),\ \xi''(t)\right] = \sigma^2\left[1-\rho(\theta)^2 - \frac{\rho'(\theta)^2}{\left[-\rho''(0)\right]} - \frac{\left[\rho''(\theta)-\rho(\theta)\rho''(0)\right]^2}{\rho^{(4)}(0)-\rho''(0)^2}\right] \quad (15)$$

As we can see from expressions Eq. 10-15, the forecast algorithms and prediction dispersion are dependent not only on random process correlation properties. Expressions Eq. 13-15 demonstrate how forecast evaluation variance goes down with an increase in the size of the sample used for prediction. It may be observed that expressions Eq. 10, 13 are widely known from literature. Expressions Eq. 13-15 can be obtained if we consider prediction ξ(t+θ/t) as a parameter of a normal stochastic process and find Fischer information for it proceeding from the resulting sample of a normal random process. Let us now write normal two-dimensional probability density for values ξ(t), ξ(t+θ) as follows:

$$f\left[\xi(t),\ \xi(t+\theta)\right] = \frac{1}{2\pi|R|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\left[\xi(t),\ \xi(t+\theta)\right]\cdot\|R\|^{-1}\left[\xi(t),\ \xi(t+\theta)\right]^T\right] \quad (16)$$

where |R|-determinant of correlation matrix ||R||, ||R||-an inverse matrix for correlation matrix ||R||:

$$\|R\| = \sigma^2 \begin{Vmatrix} 1 & \rho(\theta) \\ \rho(\theta) & 1 \end{Vmatrix} \qquad (17)$$

Further on, by finding fischer information for a sample containing one element ξ(t) relative to parameter ξ(t+θ), we have:

$$I\left[\xi(t+\theta)\right] = \mu\left[\frac{\partial^2}{\partial^2\xi(t+\theta)} \text{In} \left\{f\left[\xi(t),\, \xi(t+\theta)\right]\right\}\right] = \frac{1}{\sigma^2\left[1-\rho(\theta)^2\right]} \qquad (18)$$

where μ(x) mathematical expectation of value x. By inverting the Fischer information, we will have forecast evaluation variance ξ(t+θ) as per rao-cramer inequality and according to expression Eq. 13. In a similar way, by writing normal co-density of probability for random values ξ(t), ξ'(t), ξ(t+θ) and having in mind that matrix ||R|| for these values can be found as:

$$\|R\| = \sigma^2 \begin{Vmatrix} 1 & \rho(\theta) & \rho'(\theta) \\ \rho(\theta) & 1 & 0 \\ \rho'(\theta) & 0 & 1 \end{Vmatrix} \qquad (19)$$

we obtain Fischer information in sample ||Y|| = [ξ(t), ξ'(t)] in relation to parameter ξ(t+θ) in the form that follows:

$$I\left[\xi(t+\theta)\right] = \mu\left[\frac{\partial^2}{\partial^2\xi(t+\theta)} \text{In}\left\{f\left[\xi(t),\, \xi'(t),\, \xi(t+\theta)\right]\right\}\right] = \frac{1}{\sigma^2\left[1-\rho(\theta)^2 - \frac{\rho'(\theta)^2}{-\rho''(0)}\right]} \qquad (20)$$

By inverting the Fischer information, we obtain forecast evaluation variance ξ(t+θ) as per Rao-cramer inequality and in accordance with expression Eq. 14. Properly speaking, this only confirms the well-known fact that linear treatment is optimal for a normal random process.

Figure 1 shows dependence of normalized dispersion of forecast evaluation $l = \sigma^2 \lfloor \xi(t+0/t)/\xi(t), ..., \xi^{(n)}(t) \rfloor/\sigma^2$ for a random process with correlation function $R(\tau) = \sigma^2 \rho(\tau) = \sigma^2 \rho(\tau) = \sigma^2 \sin(\pi\pi)/(\pi\pi)$on an increase in the size of sample ||Y|| from 1 up to 5 (curves 1-5, respectively). Thus, for example, curve 1 corresponds to a sample of one random process count, curve 2 corresponds to a sample of the value of the random process and its first derivative and so, forth while curve 5 corresponds to a sample from a count of the random process and the first four derivatives.

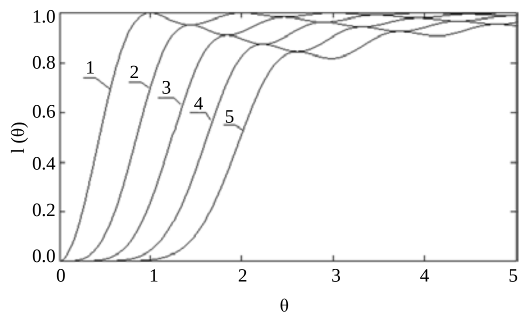

Fig. 1: Forecast dispersion $l = \sigma^2[\xi(t+\theta/t)/\xi(t), ..., \xi^{(n)}(t)]/\sigma^2$

Process $R(\tau) = \sigma^2 \sin(\pi\pi)/(\pi\pi)$ with a correlation function has finite spectral density and efficient width of random process spectral density Δf = 1 Hz which allows us to re-calculate-using a simple scaling procedure-results for the processes with another spectral density width for other forecast look-ahead times. A process with such a correlation function is widely used to illustrate random process overshoot theory. It is a linearly singular (Rozanov, 1990) or "degenerate" process and can be reconstructed by way of lineal transformation along the whole time axis. We now shall give the condition for linear singularity of a random process. If spectral density of a random process becomes zero on a positive measure set or if the following condition is met:

$$\int_{-\infty}^{\infty} \frac{\log S(\omega)}{1+\omega^2} d\omega = -\infty \qquad (21)$$

where, S(ω)-spectral density of a random process, then stochastic processes with spectral density S(ω) are singular (degenerated). Such processes with finite spectral density are believed to be farther from physical reality then, for instance, linearly regular processes. The condition of linear regularity of a random process is presented in the following form:

$$\int_{-\infty}^{\infty} \frac{\log S(\omega)}{1+\omega^2} d\omega \succ -\infty \qquad (22)$$

We can prove that linearly regular properties belong, for example, to random finitely differentiable processes. Correlation functions of such processes are shown by Tikhonov and Khimenko (1987). It seems reasonable to compare the efficiency of prediction filters based on a sample of a random process and its derivatives, to that using a sample of previous values of a random process as is the case of transversal filtering. The indicator in question is relative efficiency factor e(Δt) for samples ||Y|| and ||X|| of the same size:

$$e(\Delta t) = \frac{\sigma^2\left[\xi(t+\theta)/\|Y\|\right]}{\sigma^2\left[\xi(t+\theta)/\|X\|\right]} \qquad (23)$$

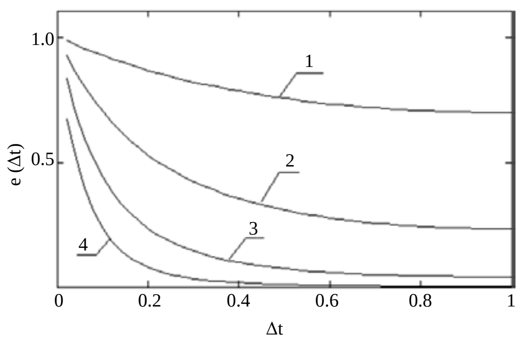

Fig. 2: Relative forecast efficiency at $\rho(\tau) = \sin(\pi\tau)/(\pi\tau)$

In Fig. 2, we can see graph $e(\Delta t)$ describing a random process with a normalized correlation function $\rho(\tau) = \sin(\pi\pi)/(\pi\pi)$ for a chosen forecast look-ahead time $\theta = 1$. In Fig. 2 curves 1-4 correspond to dependence $e(\Delta t)$ for different types of sample. Note that graph plotting was based on the same sample size of $||Y||$ and $||X||$, i.e., sample $||Y|| = [\xi(t), \xi'(t)]$ was compared with c sample $||X|| = [\xi(t), \xi(t-\Delta t)]$ and so, forth. Curve 1 corresponds to $e(\Delta t)$ for samples $||Y|| = ||\xi'(t)||$, $||X|| = ||\xi(t), \xi(t-\Delta t)||$, curve 2 corresponds to $e(\Delta t)$ for $||Y|| = ||\xi(t), \xi'(t), \xi''(t)||$, $||X|| = ||\xi(t), \xi(t-\Delta t), \xi(t-2\Delta t)||$, curve 3 corresponds to $e(\Delta t)$ for $||Y|| = ||\xi(t), \xi'(t), \xi''(t), \xi^{(3)}(t)||$, $||X|| = ||\xi(t), \xi(t-\Delta t), \xi(t-2\Delta t), \xi(t-3\Delta t)||$ and curve 4 corresponds to $e(\Delta t)$ for $||Y|| = ||\xi(t), \xi'(t), \xi''(t), \xi^{(3)}(t), \xi^{(4)}(t)||$, $||X|| = ||\xi(t), \xi(t-\Delta t), \xi(t-2\Delta t), \xi(t-3\Delta t), \xi(t-4\Delta t)||$. From graphs in Fig. 2 it is obvious that sample $||Y|| = [\xi(t), \xi'(t), ..., \xi^{(n)}(t)]$ is more efficient than sample $||X|| = ||\xi(t), \xi(t-\Delta t), ..., \xi(t-n\Delta t)||$ of an adequate dimension, particularly, so if both the dimension and time interval are increased. It can be demonstrated that at $\Delta t \to 0$ the efficiency of both samples is similar as there is similar amount of information in sample $||X|| = ||\xi(t), ..., \xi(t-\Delta t)||$ as in sample $||Y|| = [\xi(t), ..., \xi^{(n)}(t)]$ and prediction algorithms for both do converge.

RESULTS AND DISCUSSION

Wiener-Hopf linear filtering allows us to forecast a random process relying on a sample from its previous values (transversal filtering) or on a sample from the values of the random process and its derivatives. In the latter case, the correlation matrix of a random process is sparse which may have significance for elaborating adaptive filtering algorithms when the correlation matrix is being evaluated and the algorithm of direct matrix inversion used. If the time between random process counts is reduced, both algorithms and their efficiencies converge. Such prediction algorithms are optimal, first and foremost, for a normal random process; however, they can also be used to forecast stochastic processes with different probability densities.

REFERENCES

Faulkenberry, L., 1982. An Introduction to Operational Amplifiers with Linear IC Applications. Jon Wiley & Sons, New York, USA.,.

Rozanov, Y.A., 1990. [Stationary Random Processes]. Nauka Publishing, Moscow, Russia, (In Russian).

Tikhonov, V.I. and V.I. Khimenko, 1987. [Random Process Overshoots]. Nauka Publishing, Moscow, Russia, (In Russian).

Widrow, B.W. and S.D. Stearns, 1985. Adaptive Signal Processing. Prentice-Hall, New Jersey, USA..