The rôle of past and future in estimation and the reversibility of stochastic processes

Yongxin Chen, Johan Karlsson, and Tryphon Georgiou

***Abstract*— The purpose of this paper is to elucidate a dichotomy between past and future in prediction of multivariate time-series. More specifically, vector-valued gaussian stochastic processes may be deterministic in one time-direction and not the other. This fact, which is absent in scalar-valued processes, is deeply rooted in the geometry of the shift-operator. The exposition and the examples we discuss are based on the work of Douglas, Shapiro and Shields on cyclic vectors of the backward shift and relate to classical ideas going back to Wiener and Komogoroff. The paper builds on examples and the goal is to provide insight to a control engineering audience.**

I. INTRODUCTION

The variance of the one-step ahead prediction error for a scalar, stationary, discrete-time stochastic processes is given by the well-known formula

$$\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log(\Phi(\theta))d\theta\right\} \tag{1}$$

due to G. Szegö [5]. This formula expresses the error variance as the geometric mean of the power spectral density $\Phi(\theta)$ of the stochastic process. Past and future contain the same information about the present and the identical same formula provides the variance of the "postdiction" error where the present is estimated from future values. This is rather evident since (1) contains no manifestation of the time arrow. There is no such formula for the covariance matrix of the prediction or the postiction error for multivariable processes. The closest to such a formula was given by Wiener and Masani formula [11] expressing the determinant of the error covariance Ω in terms of the determinent of the power spectrum,

$$\det(\Omega) = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log(\det(\Phi(\theta)))d\theta\right\}. \tag{2}$$

Subtly, this formula leaves out the possibility of a dichotomy between past and future, which is indeed the case. This issue has been noted in classical works on prediction theory where it is pointed out that the information contained in the remote past and the information contained in the remote future may differ (see e.g., the forthcoming treatise by Lindquist and Picci [6]). Our goal at present is to highlight and elucidate this phenomenon by way of examples that are intuitively clear to an engineering audience. Prediction theory overlaps with the theory of analytic functions on the disc and of the shift operator. The exposition is relies on work of Douglas, Shapiro and Shields [2] who obtained a characterization of cyclic vectors of the "backward shift."

The issue of the time-arrow and how this is embedded in the statistics of sample-paths has been raised in the physics literature (see [9]). Typically it is thought that the time-arrow and "nonlinearities" are revealed by considering several-point correlations and higher order statistics. However, it is surprising to most that the time-arrow impacts second-order statistics and properties of stationary Gaussian processes as well. This is often missed (cf. [1], [10]) since it is exclusively a phenomenon of vector-valued processes. The point can be amply made using a simple moving average process constructed in such a way so that the prediction error differs substantially in the two time-directions (see Section III). A limit case for processes with infinite memory is when the stochastic process is not regular, and then, the process is deterministic in one of the two time-directions but not in the other. Examples include processes generated by filters whose transfer functions are cyclic with respect to the backward shift, or by a symmetric in time situation, where the process is generated by suitably a-causal filter and yet it is predictable from the infinite remote past.

The importance of the time-arrow and its manifestation in engineering and physics is hardly a new issue, yet it is one that is perhaps one of the least understood.

Supported by Swedish Research Council, National Science Foundation Grant No. 1027696, Air Force Office of Scientific Research under Grant FA9550-12-1-0319, and the Vincentine Hermes-Luh Endowment.

Y. Chen and T.T. Georgiou are with the Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA; {chen2468,tryphon}@ece.umn.edu. J. Karlsson is with the Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden; johan.karlsson@math.kth.se.

Thermodynamics and the paradox of the apparent directionality of physics originating in physical laws that are time-symmetric has not gone unnoticed; Feynman states that there is a fundamental law which says, that "uxels only make wuxels and not vice versa," but we have not found this yet. The reversibility or lack thereof of dynamical models in stochastics and control (see, e.g., [4], [8], [3]) is often puzzling, and perhaps, the time-arrow plays a fundamental role in models of engineering systems at a more basic level as well.

In Section II we review basic concepts describing the relation between stochastic time series and analytic functions. We also define cyclic vectors and present results from [2]. In Section III we present an example with a moving average process and we discuss the corresponding predictor and postdictor. In Section IV we present and example that is stochastic in one time-direction and deterministic in the other. Furthermore we describe a class of stochastic processes that have this property.

II. Preliminaries

The forward shift U is a linear operator on the Hilbert space $H_2 = H_2(\mathbb{D})$ defined as $Uf(z) = zf(z)$. We will often identify H_2 and $l_2(\mathbb{N})$, since they are isometric, and thus write

$$U : (a_0, a_1, a_2, \ldots) \to (0, a_0, a_1, \ldots).$$

The backward shift U^* is the adjoint operator of U. On H_2 we have $U^*f(z) = (f(z) - f(0))/z$, and accordingly we may write

$$U^* : (a_0, a_1, a_2, \ldots) \to (a_1, a_2, a_3, \ldots).$$

Cyclic vectors of an operator A are those vectors f such that the closure of $\{A^n f : n \geq 0\}$ spans the whole space; when f is not a cyclic vector (non-cyclic), the closure of the span is a proper invariant subspace for A. As is well known, $f \in H_2$ is cyclic for U if and only if f is an outer function. When this is not the case, f lies in the some closed invariant subspace for U, that is, subspace of the form φH_2 for some inner function φ. The invariant subspace for U^* is of the form $(\varphi H_2)^\perp$, therefore f fails to be cyclic for U^* if and only if it lies in one of the spaces $(\varphi H_2)^\perp$ for some inner function φ. This is not a property that can be easily checked! However, a more transparent condition for a function to be U^*-non-cyclic is given by the following theorem of Douglas-Shapiro-Shields.

Theorem 1 ([2]): A necessary and sufficient condition that a function f in H_2 be U^*-non-cyclic is that there exist a pair of inner functions φ and ψ such that

$$\frac{f}{\bar{f}} = \frac{\varphi}{\psi} \quad \text{almost everywhere on } \partial\mathbb{D}.$$

There are several easy but still quite interesting and intriguing properties of U^*-cyclic functions, e.g., a function is U^* cyclic if and only if its outer factor is. If f is U^* cyclic and g is non-cyclic, then $f + g$, fg and f/g are all cyclic as long as they are in H_2.

The connection and correspondence between function theory on the unit disc and discrete-time, stationary stochastic processes is well known and we will make extensive use of –our favorite concise reference is [5, Chapter 10]. The basis is the standard Kolmogoroff isomorphism between the linear space $L(x)$ generated by a second-order stochastic process $\{x_k : -\infty < k < \infty\}$ and functions on L_2 (on the unit circle). Throughout, in this correspondence, we follow the mathematical convention where U (equivalently, multiplication by z or $e^{i\theta}$) corresponds to unit time-delay for the corresponding process. Thus, in the usual slight abuse of notation, $z : x_k \mapsto x_{k-1}$ is the "delay operator" which is opposite to the way this is used in signal processing literature (where z^{-1} is often used to denote delay).

III. Comparison of predictor/postdictor error for an MA process

It is often suggested that for Gaussian stationary processes the time direction does not have an impact on the error variance (cf. [1], [10]); as noted earlier, this is not so for multivariable processes. Herein, we illustrate this fact with an example of a moving average two-variate process

$$\begin{aligned} x_k &= w_k + \alpha w_{k-1} \\ y_k &= w_k. \end{aligned}$$

The process $\{w_k \mid k \in \mathbb{Z}\}$ is taken to be gaussian, zero-mean, unit-variance and white, i.e., $E\{w_k \bar{w}_k\} = 1$, and $E\{w_k \bar{w}_\ell\} = 0$ for $k \neq \ell$. Let

$$\xi_k := \begin{bmatrix} x_k \\ y_k \end{bmatrix}$$

and consider the one-step ahead prediction problem to minimize the matrix error variance

$$E\{(\xi_k - \hat{\xi}_{k|\text{past}})(\xi_k - \hat{\xi}_{k|\text{past}})^*\}$$

in the positive-semidefinite sense, where $\hat{\xi}_{k|\text{past}}$ is a function of past measurements $x_{k-1}, x_{k-2}, \ldots,$ and $y_{k-1}, y_{k-2}, \ldots$. Since $w_k \perp x_{k-\ell}, y_{k-\ell}$ for $\ell > 0$, the solution is easily seen to be

$$\hat{\xi}_{k|\text{past}} = \begin{pmatrix} \hat{x}_k \\ \hat{y}_k \end{pmatrix} = \begin{pmatrix} \alpha y_{k-1} \\ 0 \end{pmatrix}$$

with a corresponding (forward) error variance

$$\min_{\hat{\xi}_{k|\text{past}}} E\{(\xi_k - \hat{\xi}_{k|\text{past}})(\xi_k - \hat{\xi}_{k|\text{past}})^*\} =: \Omega_{\text{f}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Likewise, since w_{k-1} is orthogonal to future measurements x_{k+1}, $x_{k+2}, \ldots$, $y_{k+1}, \ldots$, while

$$x_{k+1} - y_{k+1} = \alpha w_k,$$

we can write

$$\begin{aligned} x_k &= (x_{k+1} - y_{k+1})/\alpha + \alpha w_{k-1} \\ y_k &= (x_{k+1} - y_{k+1})/\alpha. \end{aligned}$$

The optimal estimator for x_k, y_k given future values is

$$\hat{\xi}_{k|\text{future}} = \begin{pmatrix} \hat{x}_k \\ \hat{y}_k \end{pmatrix} = \begin{pmatrix} (x_{k+1} - y_{k+1})/\alpha \\ (x_{k+1} - y_{k+1})/\alpha \end{pmatrix}$$

with corresponding minimal (backward) error variance

$$\begin{aligned} \min_{\hat{\xi}_{k|\text{future}}} \quad & E\{(\xi_k - \hat{\xi}_{k|\text{future}})(\xi_k - \hat{\xi}_{k|\text{future}})^*\} \\ & =: \Omega_\text{b} = \begin{pmatrix} \alpha^2 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

The prediction problem is clearly not symmetric with respect to time, yet $\det \Omega_\text{f} = \det \Omega_\text{b} = 0$ in agreement with the Wiener-Masani formula.

While the above example is sufficient to underscore a certain dichotomy, the forward and reversed processes have similar realizations (cf. [4]). Indeed, we can easily see that

$$\begin{aligned} x_k &= \alpha \bar{w}_k + \bar{w}_{k+1} \\ y_k &= \bar{w}_{k+1}, \end{aligned}$$

is a realization for the backward process, where $\bar{w}_k$ is a gaussian white noise process. The forward and backward realizations can be derived and correspond to the left and the right analytic factors

$$\begin{aligned} \Phi(z) &= \begin{pmatrix} 1 + \alpha z \\ 1 \end{pmatrix} (1 + \alpha z^{-1}, 1) \\ &= \begin{pmatrix} z^{-1} + \alpha \\ z^{-1} \end{pmatrix} (z + \alpha, z) \qquad (3) \end{aligned}$$

of the power spectrum $\Phi(z)$. It is possible to go one step further and construct examples where this factorization is not possible and, then, in one time-direction the process is completely deterministic.

IV. Non-reversible stochastic process

The following example underscores the situation where the power spectrum does not admit one of the two analytic factorizations and the underlying process is completely deterministic in one of the time-directions and not in the other. The stochastic process we consider is generated by

$$\begin{aligned} x_k &= w_k + \sum_{\ell=1}^{\infty} \frac{1}{1+\ell} w_{k-\ell}, \\ y_k &= w_k. \end{aligned}$$

The modeling filter

$$g(z) = \sum_{\ell=0}^{\infty} \frac{1}{1+\ell} z^\ell$$

for the x_k component has as impulse response the harmonic series. Interestingly, while this is not a stable system in an input-output sense, yet, when driven by a white noise process, it generates a well-defined L_2-process. Further, the function $g(z)$ is U^*-cyclic ([2]) and, as we will see, a direct consequence is that the process is completely deterministic in the backward time-direction.

Since $w_k \perp x_{k-\ell}, y_{k-\ell}$ for $\ell > 0$, the optimal predictor is given by

$$\hat{\xi}_{k|\text{past}} = \begin{pmatrix} \hat{x}_k \\ \hat{y}_k \end{pmatrix} = \begin{pmatrix} \sum_{\ell=1}^{\infty} \frac{1}{1+\ell} y_{k-\ell} \\ 0 \end{pmatrix}$$

with a corresponding (forward) error variance

$$\inf_{\hat{\xi}_{k|\text{past}}} E\{(\xi_k - \hat{\xi}_{k|\text{past}})(\xi_k - \hat{\xi}_{k|\text{past}})^*\} =: \Omega_\text{f} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Reversing time-direction, we now want to estimate x_0, y_0 given future observations, x_ℓ, y_ℓ, $\ell = 1, 2, \ldots$. Since $w_\ell = y_\ell$, equivalently, we want to estimate x_0, y_0 given

$$\begin{aligned} \tilde{x}_k &:= x_k - \sum_{\ell=1}^{k} \frac{1}{k - \ell + 1} y_\ell \\ &= \sum_{\ell=0}^{\infty} \frac{1}{k+\ell} w_{-\ell} \end{aligned}$$

and y_k, for $k = 1, 2, \ldots$. Since, now, $\overline{\text{span}}_{k>0}\{y_k\}$ is orthogonal to x_0, y_0, and $\overline{\text{span}}_{k>0}\{\tilde{x}_k\}$, this is equivalent to estimating x_0, y_0 based only on $\tilde{x}_k$ for $k = 1, 2, \ldots$. We now note that

$$\sum_{\ell=0}^{\infty} \frac{1}{k+\ell} z^k = (U^*)^k g(z).$$

Also note that $U^* g(z)$ is U^*-cyclic since $g(z)$ is U^*-cyclic.[1] So we have, $\overline{\text{span}}_{k\geq 1}\{(U^*)^k g(z)\} = H_2$, and hence

$$\overline{\text{span}}_{k\geq 1}\{\tilde{x}_k\} = \overline{\text{span}}_{k\leq 0}\{w_k\} \ni x_0, y_0.$$

The infimum of the backward error variance is therefore

$$\begin{aligned} \inf_{\hat{\xi}_{k|\text{future}}} \quad & E\{(\xi_k - \hat{\xi}_{k|\text{future}})(\xi_k - \hat{\xi}_{k|\text{future}})^*\} \\ & =: \Omega_\text{b} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

[1] This follows since $g(z)$ is cyclic, $g(0)$ and z are non-cyclic, and $U^* g(z) = (g(z) - g(0))/z \in H_2$.

and the time series is uniquely determined by the infinite future (c.f., [5], [6]). Building on this example and using [2] we derive the following result.

Theorem 2: Consider the stochastic processes

$$\begin{aligned} x_k &= \sum_{\ell=0}^{\infty} g_\ell w_{k-\ell}, \\ y_k &= \sum_{\ell=0}^{\infty} h_\ell w_{k-\ell} \end{aligned}$$

where $g(z) = \sum_{\ell=0}^{\infty} g_\ell z^\ell$ is cyclic and $h(z) = \sum_{\ell=0}^{\infty} h_\ell z^\ell \neq 0$ is non-cyclic. Then the backward process is deterministic.

V. Concluding remarks

While the moving average process in Section III admits a spectral factorization for the backward process (3), there is no such factorization for the non-reversible process in Section IV. This may be viewed in the context of decompositions for stochastic processes that are Gaussian, zero mean, and stationary. Namely, the Hilbert space generated by any such process may be decomposed as

$$\mathbf{H}(\xi_k) = \mathbf{H}_{-\infty}(\xi_k) \oplus \mathbf{H}(w_k),$$

in terms of the remote past and the driving noise, namely,

$$\begin{aligned} \mathbf{H}(\xi_k) &= \overline{\text{span}}_{k\in\mathbb{Z}}\{\xi_k\} \\ \mathbf{H}_{-\infty}(\xi_k) &= \cap_{t\in\mathbb{Z}}\overline{\text{span}}_{k\leq t}\{\xi_k\}, \quad \text{and} \\ \mathbf{H}(w_k) &= \overline{\text{span}}_{k\in\mathbb{Z}}\{w_k\}, \end{aligned}$$

and similarly in terms of the remote future and the Hilbert space generated in the backward direction by a driving noise

$$\mathbf{H}(\xi_k) = \mathbf{H}_{+\infty}(\xi_k) \oplus \mathbf{H}(\bar{w}_k)$$

(see [6] for details). The process is reversible if the remote past and the remote future coincide. Here we have considered a concrete example of a non-reversible process where the remote past is trivial while the remote future spans the entire process.

The essence of this example (Section IV and cf. [6]), is that the power spectrum of $\{(x_k, y_k)^T\}$, being

$$\begin{pmatrix} g(z) \\ 1 \end{pmatrix} \begin{pmatrix} g(z)^* & 1 \end{pmatrix}$$

fails to have a co-analytic spectral factorization. This can be shown using Theorem 1 (see also [6]). It also fails to satisfy condition 3 of Theorem 2 in [7]. This absence of co-analytic factorization renders the backward process deterministic.

Finally, the dependence between of x_0 and $\tilde{x}_k$ on w_k for $k \in \mathbb{Z}$ is expressed as follows,

$$\begin{pmatrix} x_0 \\ \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\ \vdots & \vdots & \vdots & \\ \frac{1}{n+1} & \frac{1}{n+2} & \frac{1}{n+3} & \cdots \end{pmatrix} \begin{pmatrix} w_0 \\ w_{-1} \\ w_{-2} \\ \vdots \end{pmatrix}.$$

It would be interesting to establish the U^*-cyclicity of $g(z)$ directly based on spectral properties of the Hilbert matrix that expresses this relationship.

We will further explore these issues in order to gain understanding on how prediction and postdiction in stochastic processes relate to time directionality and causality.

VI. Acknowledgment

We would like to thank Anders Lindquist for discussions and insights on the subject.

References

[1] C. Diksayl, J.C. van Houwelingen, F. Takens, and J. DeGoede ”Reversibility as a criterion for discriminating time series,” Physics Letters A 201, 221-228, May 1995.

[2] R. G. Douglas, H. S. Shapiro, and A. L. Shields. ”Cyclic vectors and invariant subspaces for the backward shift operator.” Annales de l’institut Fourier. Vol. 20. No. 1. Institut Fourier, 1970.

[3] T. T. Georgiou and M. C. Smith, ”Feedback control and the arrow of time.” International Journal of Control 83.7 (2010): 1325-1338.

[4] T. T. Georgiou and A. G. Lindquist, ”On time-reversibility of linear stochastic models.” arXiv preprint arXiv:1309.0165 (2013).

[5] U. Grenander and G. Szegö, Toeplitz Forms and their Applications, Chelsea, 1958.

[6] A. G. Lindquist and G. Picci, ”Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification,” a treatise on stochastic realization theory, in preparation/circulation, version 2013.

[7] Y. A. Rozanov, ”Spectral properties of multivariate stationary processes and boundary properties of analytic matrices,” Theory of probability and its applications, Vol. V, No. 4, (1960).

[8] H. Sandberg, J. Delvenne, and J.C. Doyle. ”On lossless approximations, the fluctuation-dissipation theorem, and limitations of measurements.” Automatic Control, IEEE Transactions on 56.2 (2011): 293-308.

[9] Tong, Howell. Non-linear time series: a dynamical system approach. Oxford University Press, 1990.

[10] G. Weiss, ”Time-reversibility of linear stochastic processes.” Journal of Applied Probability (1975): 831-836.

[11] N. Wiener and P. Masani. ”The prediction theory of multivariate stochastic processes.” Acta Mathematica 98.1 (1957): 111-150.