

The Role of the Time-Arrow in Mean-Square Estimation of Stochastic Processes

Yongxin Chen, Johan Karlsson, and Tryphon T. Georgiou

Abstract—The purpose of this letter is to point out a certain dichotomy between the information that the past and future values of a *multivariate* stochastic process carry about the present. More specifically, vector-valued, second-order stochastic processes may be deterministic in one time-direction but not in the other. This phenomenon, 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 Kolmogorov. We focus on rank-one stochastic processes for which we obtain an explicit characterization of all regular processes that are deterministic in the reverse time-direction. This letter builds on examples and the goal is to provide insights to a control engineering audience with interests in estimation theory and modeling of time-series.

Index Terms—Linear stochastic systems, estimation.

I. INTRODUCTION

THE VARIANCE of the error in predicting, one-step-ahead, the values of a *scalar*, second-order, stationary, discrete-time stochastic process is given by a well-known formula due to Grenander and Szegő [1] as the geometric mean

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

of its power spectral density $\Phi(\theta)$. Reversal of the time direction does not impact this formula. Thus, the past and the future of the process contain the same information about the present, and the identical same formula provides the variance of the “postdiction” error when the present is estimated

from future values. In contrast, for multivariable processes, the matrix-covariance of the prediction error requires spectral factorization and cannot be expressed explicitly in terms of the matrix-valued power spectral density in a similar manner. The closest to such a formula, given by Wiener and Masani [2, p. 145, Main Theorem I], expresses the determinant of the error matrix-covariance, herein denoted by Ω , as the geometric mean of the determinant of the power spectrum,

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

In a subtle way, when $\det(\Omega) = 0$, this formula leaves out the possibility of a dichotomy between past and future, and as it turns out this is indeed the case. More precisely, it is perfectly possible for a (multivariable) stationary Gaussian stochastic process to be purely deterministic in one time-direction but not in the other. This issue has been noted in classical works in prediction theory where it has been pointed out that the information contained in the remote past and the information contained in the remote future may differ, see [3, Sec. 4.5]. Thus, one objective of the present work is to highlight and elucidate this phenomenon with examples that are intuitively clear to an engineering audience (Sections III and IV). By expanding on the insight gained, we provide a characterization of rank-1 regular processes that are completely deterministic in one time direction (Sec. V).

Broadly speaking, the manifestation of the time-arrow in engineering and physics is hardly a new issue, yet it is one that is not well understood. The paradox of the apparent directionality of physics originating in physical laws that are time-symmetric is a key conundrum; Feynman states that there is a fundamental law which says, that “uxels only make wuxels and not vice versa,” but that we have not found this yet. Thus, the time reversibility of physical models, as well as the lack of, remain of great scientific interest, see [4], [5]. In a similar vein, we expect that issues related to the time-arrow will draw increasing attention in modeling of control systems as well, see [6].

Turning to time-series, the possible ways in which the time-arrow is encoded in the statistics have also been studied in the physics literature as well, see [7]. It is widely thought that the time-direction and “nonlinearities” are revealed by considering several-point correlations and higher order statistics. While this may be so at times, it is surprising to most that *the time-arrow may already be clearly discerned in second-order stationary processes as well, in that their predictability*

Manuscript received June 22, 2017; accepted August 4, 2017. Date of publication August 17, 2017; date of current version August 30, 2017. This work was supported in part by Swedish Research Council, in part by NSF under Grant ECCS-1509387, and in part by AFOSR under Grant FA9550-17-1-0435 and Grant FA9550-15-1-0045. A preliminary version of this work was presented in [17]. Recommended by Senior Editor G. Yin. (Corresponding author: Tryphon T. Georgiou.)

Y. Chen is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: yongchen@iastate.edu).

J. Karlsson is with the Department of Mathematics, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: johan.karlsson@math.kth.se).

T. T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California at Irvine, Irvine, CA 92697 USA (e-mail: tryphon@uci.edu).

Digital Object Identifier 10.1109/LCSYS.2017.2740957

properties may dramatically differ depending on the time-direction. The reason that this observation is often missed (see [8], [9]) may be due to the fact that it is exclusively a phenomenon of vector-valued processes. In fact, it is a common misconception [10] that “temporal irreversibility is important because it excludes Gaussian linear dynamics and static transformations... from the set of possible generating processes.” Thus, one of the aims of this letter is to explain why this is not so. As noted, we provide an example of a vector-valued moving-average process constructed so that the prediction error differs substantially in the two time-directions (Section III). A limit case of a stochastic process with infinite memory allows it to be deterministic in one of the two time-directions but not in the other (Section IV).

Prediction theory of second-order processes overlaps with that of analytic functions on the unit disc and the shift operator. Thus, the exposition and technical results of this letter rely heavily on this connection and on the work of Douglas *et al.* [11] who obtained a characterization of cyclic vectors of the “backward shift.” Our analysis and examples include processes generated by filters whose transfer functions are cyclic with respect to the backward shift, or in a time-symmetric situation, processes generated by suitable acausal filters that are predictable from the infinite remote past. Besides explaining the dichotomy between past and future, and on how this relates to factorizability of the power spectrum [12], [13], we also study regular rank-one processes and explicitly characterize all such processes that are deterministic in the reverse-time direction.

II. NOTATIONS AND PRELIMINARIES

A. Function Theory

The notation used in this letter is now briefly defined as it is standard. We denote by $\mathbb{R}, \mathbb{C}, \mathbb{Z}$ real numbers, complex numbers, and integers, respectively, by $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ the unit disc and by $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ the unit circle on the complex plane. We denote by $L_2(\mathbb{T})$ the Hilbert space of square-integrable functions on $\mathbb{T}$, by $H_2 \subset L_2(\mathbb{T})$ the Hardy space of functions whose negative Fourier coefficients vanish, by $\ell_2(\mathbb{Z}_+)$ the space of square summable sequences on the nonnegative integers $\mathbb{Z}_+$, and by $\|\cdot\|_2$ the norm in the respective spaces. As is well known, H_2 can be identified with the space of analytic functions $f(z)$ with $z \in \mathbb{D}$ having square integrable radial limits. In fact, the correspondence

$$f(z) = \sum_{k \geq 0} f_k z^k \mapsto (f_0, f_1, \dots)$$

is a Hilbert space isomorphism between H_2 and $\ell_2(\mathbb{Z}_+)$. Likewise, $L_2(\mathbb{T})$ and $\ell_2(\mathbb{Z})$ are Hilbert isomorphic. The orthogonal complement of H_2 in $L_2(\mathbb{T})$ is denoted by $H_2^\perp = \{\sum_{k < 0} f_k z^k \in L_2(\mathbb{T})\}$ while H_2^- is a short for $zH_2^\perp = \{\sum_{k \leq 0} f_k z^k \in L_2(\mathbb{T})\}$. We use A^* to denote the adjoint of the operator A . Conformably, if A is a matrix and $f(z)$ a function of z , A^* is the conjugate transpose and $f^*(z) = \overline{f(\bar{z}^{-1})}$ where $\bar{\cdot}$ denotes complex conjugation.

The forward shift U is a linear operator on H_2 defined by $Uf(z) = zf(z)$. We use the same symbol for the shift in $\ell_2(\mathbb{Z}_+)$:

$$U : (f_0, f_1, f_2, \dots) \rightarrow (0, f_0, f_1, \dots).$$

The backward shift U^* is the adjoint operator of U [14]. On H_2 , it is $U^*f(z) = (f(z) - f(0))/z$, and on $\ell_2(\mathbb{Z}_+)$

$$U^* : (f_0, f_1, f_2, \dots) \rightarrow (f_1, f_2, f_3, \dots).$$

A vector f (or, a function if we are dealing with H_2) is a cyclic vector of A if the closure of the span of $\{A^n f : n \geq 0\}$ is the complete space; if f is not cyclic then the closure of the span is a proper A -invariant subspace. Cyclic vectors of U are precisely the outer functions in H_2 [14]; in the engineering literature these are referred to as *stable and minimum phase* (or, loosely, *stable and stably invertible*). When f is not outer, it lies in a closed invariant subspace of U , i.e., one of the form φH_2 for some *inner*¹ function φ . An invariant subspace of U^* is of the form $(\varphi H_2)^\perp$. Therefore f fails to be cyclic under U^* if and only if it is orthogonal to one of the spaces φH_2 with φ inner. This is a property that is difficult to verify in general. An explicit characterization for the failure of $f \in H_2$ to be cyclic with respect to U^* is provided next (see [11]).

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

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

There are several easy but quite surprising properties of U^* cyclic functions as noted in [11], in particular, i) a function is U^* cyclic if and only if its outer factor is, and ii) 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 . Throughout the rest of this letter, “cyclic” means cyclic with respect to U^* unless otherwise stated.

The shift operator extends to a unitary operator on $L_2(\mathbb{T})$. For this we use the same symbol as that for the shift on H_2 , namely $Uf(z) = zf(z)$. Evidently, U is invertible with $U^{-1}f(z) = z^{-1}f(z)$. A characterization of simply invariant subspaces of $L_2(\mathbb{T})$ (i.e., invariant with respect only one of U^{-1} and U) will be needed and is as follows (see [15, p. 8, Th. 3]).

Proposition 1: If a subspace $M \subset L_2$ is U^{-1} invariant but not U invariant, then it has the form $M = qH_2^-$ for some unimodular² function q .

B. Second-Order Stochastic Processes

For $\{x_k \mid k \in \mathbb{Z}\}$, a zero-mean discrete time second order stochastic process, we use $\text{span}_{k \in \mathbb{Z}}\{x_k\}$ to denote the space of all finite linear combinations of elements in $\{x_k\}_{k \in \mathbb{Z}}$ — these are random variables on a suitable probability space, and we use

$$\mathbf{H}(x) := \overline{\text{span}}_{k \in \mathbb{Z}}\{x_k\}$$

to denote the closure of the span. This is a Hilbert space where, as usual, the inner product between random variables is $\langle x, y \rangle = \mathbb{E}\{x\bar{y}\}$ and $\mathbb{E}$ is the expectation operator, see [1, p. 167]. For Gaussian zero-mean random variables, orthogonality (i.e., $\mathbb{E}\{x\bar{y}\} = 0$) is equivalent to being independent.

¹A function $f(z) \in H_2$ is inner [14] if it has radial limit $\lim_{r \rightarrow 1} |f(re^{i\theta})| = 1$ for almost all $\theta \in [0, 2\pi)$. Such functions are known in the engineering literature as *all-pass*.

²A function $q \in L_2(\mathbb{T})$ is unimodular if $|q(z)| = 1$ a.e. on $\mathbb{T}$.

The correspondence between function theory on the unit disc and discrete-time, stationary stochastic processes is well known, see [1, Ch. 10] for a concise exposition. The basis of this correspondence is the *Kolmogorov isomorphism* between the linear space generated by second-order stochastic processes and functions on the unit circle. In particular, consider $\{w_k \mid k \in \mathbb{Z}\}$ to be (complex-valued) Gaussian, zero-mean, unit-variance, white noise, i.e., a stochastic process such that $\mathbb{E}\{w_k \bar{w}_k\} = 1$, and $\mathbb{E}\{w_k \bar{w}_\ell\} = 0$ for $k \neq \ell$. The map

$$\mathbf{H}(w) \rightarrow L_2(\mathbb{T}) : \sum_k f_k w_{-k} \mapsto \sum_k f_k z^k \quad (3)$$

is a Hilbert space isomorphism. For a non-white process, $\{x_k\}$, $\mathbf{H}(x)$ corresponds to squarely integrable functions on $\mathbb{T}$ with respect to a suitable spectral measure (see [1, p. 175]) but this will not be needed in the sequel.

As is evident from the above, $f(z) \in H_2$ corresponds to a random variable $\sum_k f_k w_{-k}$ which can be thought as the output

$$x_n = \sum_k f_k w_{n-k}$$

at $n = 0$, of a linear system with $\{f_k\}$ as its impulse response and input the white noise process. The shift operator U in ℓ_2 corresponds to multiplication by z or $e^{i\theta}$ in $L_2(\mathbb{T})$ and to a unit time-delay in $\mathbf{H}(w)$.

III. COMPARISON OF PREDICTOR/POSTDICTION ERROR FOR A MOVING-AVERAGE PROCESS

It is often suggested that for Gaussian stationary processes, the time direction does not have an impact on the error variance (see [8], [9]). Also, as already noted in the introduction, it is also a common misconception [10, p. 1912] to believe that temporal irreversibility excludes Gaussian linear dynamics as a possible generating mechanism. The purpose of this letter is to explain that this is not so. However, in order to produce such a counterexample of a Gaussian stochastic process that has distinctly different features in the two time directions, one needs to consider multivariable processes. We illustrate the mechanism of how this can happen by an example that we present next.

Consider the moving-average bivariate process $\xi_k := (x_k, y_k)^T$ defined by the filter equations

$$x_k = w_k + \alpha w_{k-1} \quad (4a)$$

$$y_k = w_k, \quad (4b)$$

where $\alpha \neq 0$ and the process $\{w_k \mid k \in \mathbb{Z}\}$ is complex-valued Gaussian, zero-mean, unit-variance and white. We are interested in one-step ahead linear prediction.³ Thus, we seek to minimize the (matrix) error-variance

$$\mathbb{E}\{(\xi_0 - \hat{\xi}_{0|\text{past}})(\xi_0 - \hat{\xi}_{0|\text{past}})^*\}$$

in the positive-semidefinite sense. Here, $\hat{\xi}_{0|\text{past}}$ is a function of past measurements $x_{-1}, x_{-2}, \dots$, and $y_{-1}, y_{-2}, \dots$. In the forward time-direction, since w_0 is independent of $x_{-\ell}, y_{-\ell}$

for $\ell > 0$, the solution is easily seen to be

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

with a corresponding **forward error variance**

$$\Omega_f := \min_{\hat{\xi}_{0|\text{past}}} \mathbb{E}\{(\xi_0 - \hat{\xi}_{0|\text{past}})(\xi_0 - \hat{\xi}_{0|\text{past}})^*\} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

In the reverse time direction, since $x_{k+1} - y_{k+1} = \alpha w_k$, we can write the dynamics (4) as

$$x_k = (x_{k+1} - y_{k+1})/\alpha + \alpha w_{k-1}$$

$$y_k = (x_{k+1} - y_{k+1})/\alpha.$$

Similar to the above argument for the forward time-direction, w_{-1} is orthogonal to future measurements $x_1, x_2, \dots$, and $y_1, y_2, \dots$, and hence, given future values, the optimal estimator for x_0, y_0 is

$$\hat{\xi}_{0|\text{future}} = \begin{pmatrix} \hat{x}_{0|\text{future}} \\ \hat{y}_{0|\text{future}} \end{pmatrix} = \begin{pmatrix} (x_1 - y_1)/\alpha \\ (x_1 - y_1)/\alpha \end{pmatrix}$$

with corresponding minimal backward error variance

$$\Omega_b := \min_{\hat{\xi}_{0|\text{future}}} \mathbb{E}\{(\xi_0 - \hat{\xi}_{0|\text{future}})(\xi_0 - \hat{\xi}_{0|\text{future}})^*\} = \begin{pmatrix} \alpha^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

The prediction problem is clearly not symmetric with respect to time, yet $\det \Omega_f = \det \Omega_b = 0$ in agreement with the Wiener-Masani formula [2, p. 145, Main Theorem I].

The above example is sufficient to underscore the dichotomy. The forward and reversed processes have similar realizations (see [5]). Indeed, we can easily see that

$$x_k = \alpha \tilde{w}_k + \tilde{w}_{k+1}$$

$$y_k = \tilde{w}_{k+1},$$

is a backward-in-time realization of the process, where $\tilde{w}_k$ is a standard Gaussian white-noise process. The forward and backward realizations can be derived and correspond to the left and the right analytic factors

$$\Phi(z) = \begin{pmatrix} 1 + \alpha z \\ 1 \end{pmatrix} (1 + \alpha z^{-1}, 1) = \begin{pmatrix} z^{-1} + \alpha \\ z^{-1} \end{pmatrix} (z + \alpha, z) \quad (5)$$

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

IV. A NON-REVERSIBLE STOCHASTIC PROCESS

The following example presents a case where the power spectrum **does not admit one of the two spectral factorizations**. As a consequence, the process is completely deterministic in one of the time-directions and not in the other. The stochastic process we consider is generated by

$$x_k = w_k + \sum_{\ell=1}^{\infty} \frac{1}{1+\ell} w_{k-\ell}, \quad (6a)$$

$$y_k = w_k. \quad (6b)$$

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,

³Without loss of generality we consider estimating x_0, y_0 , i.e., for $k = 0$, since all processes in this letter are stationary.

while this is not a stable system in an input-output sense, when driven by a white-noise process, it generates a well-defined stochastic process with finite variance since the harmonic series is square-summable. Further, the function $g(z)$ is cyclic [11] and, as we will see, a direct consequence is that the process is completely deterministic in the backward time-direction.

Since w_0 is orthogonal to $x_{-\ell}, y_{-\ell}$ for $\ell > 0$, the optimal predictor is given by

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

with a corresponding (forward) error variance

$$\Omega_f = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

In the reverse time-direction, we estimate x_0, y_0 given future observations, x_ℓ, y_ℓ , for $\ell > 0$. Since $w_\ell = y_\ell$, the infinite sequence of random variables $w_1, w_2, \dots$ is available for estimating x_0 and y_0 . Define $\tilde{x}_1 = x_1$, and

$$\tilde{x}_k := x_k - \sum_{\ell=0}^{k-2} \frac{1}{1+\ell} w_{k-\ell} = \sum_{\ell=k-1}^{\infty} \frac{1}{1+\ell} w_{k-\ell},$$

for $k \geq 2$. It follows that $\overline{\text{span}}_{k \geq 1} \{y_k, x_k\} = \overline{\text{span}}_{k \geq 1} \{y_k, \tilde{x}_k\}$. Now, notice that $\overline{\text{span}}_{k \geq 1} \{y_k\}$ is orthogonal to x_0, y_0, y_1 as well as $\overline{\text{span}}_{k \geq 0} \{\tilde{x}_k\}$. This is due to the fact that w_k for $k \geq 2$ is orthogonal to $x_0, y_0, y_1, \tilde{x}_1, \tilde{x}_2, \dots$. Hence the estimation problem given above is equivalent to estimating x_0, y_0 based on y_1 and $\tilde{x}_k$ for $k \geq 1$. In fact, y_1 is not needed and as we will see next, x_0, y_0 can be predicted with arbitrary precision based only on $\tilde{x}_k$ for $k \geq 1$.

The relation between $\tilde{x}_k$ and w_k for $k \in \mathbb{Z}$ can be expressed as

$$\tilde{\mathbf{x}} := \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} w_1 \\ w_0 \\ w_{-1} \\ \vdots \end{pmatrix} =: \mathcal{H} \mathbf{w}, \quad (7)$$

where $\mathcal{H}$ denotes the (infinite) Hilbert matrix, or equivalently, the representation of a Hankel operator with symbol the harmonic series. Note that the $(k+1)$ th row of $\mathcal{H}$ corresponds to the backward shifted input responses

$$\sum_{\ell=0}^{\infty} \frac{1}{k+\ell+1} z^\ell = U^{*k} g(z), \quad \text{for } k = 0, 1, \dots,$$

and since $g(z)$ is cyclic [11, Th. 2.2.3], we have $\overline{\text{span}}_{k \geq 0} \{U^{*k} g(z)\} = H_2$. Thus, linear combinations of $\tilde{x}_k$ for $k = 1, 2, \dots$ can approximate with arbitrarily small error any linear combination of $w_0, w_{-1}, \dots$ having finite norm. Hence, in particular, linear combinations of $\tilde{x}_k$ for $k = 1, 2, \dots$ can also approximate⁴ with arbitrarily small error x_0 and y_0 , since these correspond to elements in H_2 (i.e., linear combinations of $w_0, w_{-1}, \dots$ having finite norm). The infimum of the backward

error variance is therefore

$$\Omega_b = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and the time series $\{\xi_k\}$ is uniquely determined by the infinite future (see [1], [3]).

V. CHARACTERIZATION OF BACKWARD DETERMINISTIC RANK-ONE PROCESSES

An n -dimensional Gaussian stochastic process is regular if its spectrum admits a (right) analytic factorization (see [13]) and, thereby, can be represented in the form

$$\xi_k = \sum_{\ell=0}^{\infty} G_\ell w_{k-\ell}$$

where $w_k \in \mathbb{C}^m$ is a white-noise process, and the sequence $\{G_k\}_{k \geq 0} \in \ell_2^{n \times m}$ [2], [13]. Rank-one regular processes are those where the white-noise process w_k may be taken as a scalar process ($m = 1$).

Building on the example from the previous section and using results from [11], we characterize all the regular rank-one processes that are backward deterministic. We start by identifying a subclass of bivariate processes that contains the example from Section IV. Below we take $G_\ell = (g_\ell, h_\ell)^T \in \mathbb{C}^2$, that is $g_\ell, h_\ell \in \mathbb{C}$.

Theorem 2: Consider the stochastic processes

$$x_k = \sum_{\ell=0}^{\infty} g_\ell w_{k-\ell}, \quad (8a)$$

$$y_k = \sum_{\ell=0}^{\infty} h_\ell w_{k-\ell} \quad (8b)$$

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-in-time realization of the process is deterministic.

To show this, we need the following lemma.

Lemma 1: If h is a non-cyclic function, then there exists an inner function ψ such that

$$\overline{\text{span}}_{k \geq 0} (z^{-k} \psi h) \supset H_2^\perp.$$

Proof: See the Appendix. ■

Proof of Theorem 2: Let the inner function ψ be selected according to Lemma 1 so that $\overline{\text{span}}_{k \geq 0} (z^{-k} \psi h) \supset H_2^\perp$. The backward prediction error for x_0 is bounded by

$$\begin{aligned} \sqrt{\mathbb{E}(\|x_0 - \hat{x}_{0|\text{future}}\|^2)} &= \|a^* g + b^* h\|_2 = \|\psi(a^* g + b^* h)\|_2 \\ &\leq \|P_{H_2}(\psi a^* g)\|_2 + \|\psi b^* h + P_{H_2^\perp}(\psi a^* g)\|_2 \end{aligned} \quad (9)$$

where a and b are polynomials with $a(0) = 1$ and $b(0) = 0$, corresponding to the predictor $\hat{x}_{0|\text{future}} = -\sum_{\ell=1}^n (\bar{a}_\ell x_\ell + \bar{b}_\ell y_\ell)$. The inequality in (9) follows from the triangle inequality. Since g is cyclic, so is $U^*(g\psi)$, hence $\text{span}_{\ell \geq 1} U^{*\ell} \psi g$ is dense in H_2 , and therefore the first term of (9) which equals

$$\|P_{H_2}(\psi a^* g)\|_2 = \|\psi g + \sum_{\ell=1}^{\infty} \bar{a}_\ell U^{*\ell}(\psi g)\|_2,$$

can be made arbitrarily small by selecting the polynomial a properly. Since $\overline{\text{span}}_{k \geq 0} (z^{-k} \psi h) \supset H_2^\perp$, the polynomial b can

⁴Note that this approach is non-constructive and does not give an explicit expression for the linear combination.

be selected so that the second term of (9) is arbitrarily small as well.

A similar argument can be used to show that y_0 can be estimated with arbitrarily small error, by considering polynomials with $a(0) = 0$ and $b(0) = 1$ in (9). This completes the proof. ■

Following the same line of reasoning as in the proof of Theorem 2, one can in fact show that

$$\overline{\text{span}}_{k>0}\{z^{-k}g\} + \overline{\text{span}}_{k>0}\{z^{-k}h\} = L_2,$$

and therefore, the backwards prediction error is zero as a result of the correspondence in (3). Thus, in this case, the input sequence w_k , for $k \in \mathbb{Z}$, may be reconstructed arbitrarily well from the future output, x_k, y_k for $k > 0$.

As stated in Theorem 1 [11], a function $g \in H_2$ is cyclic if and only if $g/\bar{g}$ belong to $\mathcal{J}$, the set of unimodular functions that are quotients of inner functions,⁵ i.e., $\mathcal{J} = \{\varphi/\psi : \varphi, \psi \text{ inner}\}$. This result is central to our characterization of backward deterministic rank-one processes, and leads to our main result.

Theorem 3: Let $g, h \in H_2$, then the following conditions are equivalent

- (a) The system (8) is backward deterministic,
- (b) $\overline{\text{span}}_{k \geq 0}\{z^{-k}g\} + \overline{\text{span}}_{k \geq 0}\{z^{-k}h\} = L_2$,
- (c) $gh/(\bar{g}h) \notin \mathcal{J}$.

Proof: See the Appendix. ■

Note that Theorem 2 follows as a special case, from the equivalence between (a) and (c) and by using the fact that $g/\bar{g} \notin \mathcal{J}$ and $h/\bar{h} \in \mathcal{J}$ (see Theorem 1). In view of Theorem 3, we define backward deterministic processes generated by a set of functions as follows.

Definition 1: The functions $g^{(1)}, g^{(2)}, \dots, g^{(n)} \in H_2$ are called backward deterministic if

$$\sum_{j=1}^n \overline{\text{span}}_{k \geq 0}\{z^{-k}g^{(j)}\} = L_2. \quad (10)$$

As a corollary to Theorem 3 we also obtain an analogous result for general vector-valued rank-one processes.

Corollary 1: The non-zero functions $g^{(1)}, g^{(2)}, \dots, g^{(n)} \in H_2$ are backward deterministic if and only if

$$\frac{g^{(1)}\bar{g}^{(j)}}{\bar{g}^{(1)}g^{(j)}} \notin \mathcal{J} \quad (11)$$

for some $j = 2, \dots, n$.

Proof: This follows using Theorem 3. ■

VI. CONCLUSION

In this letter we pointed out a certain dichotomy in stochastic models where a process can be deterministic in one time direction and not the other. This appears counterintuitive and requires considerable mathematical sophistication to explain. Our goal in this letter has been to provide an analysis as well as insights aimed at a control audience with interests in estimation theory. In particular, we provided concrete examples of non-reversible processes where the remote past is trivial while the remote future spans the entire process. The essence of these

examples (Sections IV and V and also see [3, Sec. 4.5]) is that the power spectrum of the vector-valued stochastic process $\{\xi_k\}$,

$$\Phi(z) = \begin{pmatrix} g(z) \\ h(z) \end{pmatrix} \begin{pmatrix} g^*(z) & h^*(z) \end{pmatrix},$$

fails to have a co-analytic spectral factorization, i.e., a factorization of the form $F^*(z)F(z)$ with F analytic in $\mathbb{D}$; an equivalent statement is that the backward-in-time process is not regular [13]. This can be shown using Theorem 1. The power spectrum also fails to satisfy condition 3 of [13, Th. 2]. This absence of co-analytic factorization renders the backward-in-time realization of the process deterministic.

While the issues pointed out are quite technical, they impact in significant ways the relevance of certain models for time-series. Naturally, past and future are important in smoothing. Hence, the time asymmetry of stochastic models can be of great interest from an engineering standpoint. As noted, this asymmetry manifests itself in the absence of left or right analytic factorizations for the corresponding (necessarily irrational) power spectra. But even when factorizations exist (e.g., when the power spectrum is nonsingular at all frequencies), the limiting case where spectral factors fail to exist requires further understanding. In particular, in such cases, it is of interest to quantify the different rates that information accrues with increasing amount of data in the past or future of a particular point in time for estimation purposes (i.e., smoothing utilizing fixed window of observations). We expect that these rates relate to mixing rates for stochastic processes in corresponding time directions—a subject of independent interest.

APPENDIX

Proof of Lemma 1: Since h is non-cyclic there exists an inner function φ such that $\text{span}_{k>0}U^{*k}h = (\varphi H_2)_{H_2}^\perp$, hence

$$\begin{aligned} \overline{\text{span}}_{k>0} z^{-k}h &\subset (\varphi H_2)_{L_2}^\perp \\ \overline{\text{span}}_{k>0} z^{-k}h\bar{\varphi}z &\subset H_2^-. \end{aligned}$$

Since the left hand side is invariant with respect to z^{-1} it is on the form $\bar{\psi}H_2^-$ where ψ is inner. From Beurling-Lax theorem [15], [18] it follows that

$$\begin{aligned} \overline{\text{span}}_{k>0} z^{-k}h\bar{\varphi}z &= \bar{\psi}H_2^- \\ \overline{\text{span}}_{k>0} z^{-k}h\psi &\supset \overline{\text{span}}_{k>0} z^{-k}h\psi\bar{\varphi} = H_2^\perp. \end{aligned}$$

Proof of Main Theorem (Theorem 3): In order to prove the main theorem we will use the following lemmas.

Lemma 2: For any function $g \in H_2$, the subspace $\overline{\text{span}}_{k \geq 0}\{z^{-k}g\}$ is equal to qH_2^- , where $q = g/\bar{g}_{\text{outer}}$ and g_{outer} is the outer part of g .

Proof: Clearly $M = \overline{\text{span}}_{k \geq 0}\{z^{-k}g\} \subset L_2$ is a invariant subspace for U^{-1} while not for U , so by Proposition 1 it has the form $M = qH_2^-$ for some unimodular function q and hence $M = q \oplus z^{-1}M$. The function q is determined by the subspace up to a constant factor. We next compute one such q . Since $q \in M$, we have that $q = g\bar{f}$ for some analytic function f . We claim that one feasible f is given by

$$f = 1/g_{\text{outer}},$$

⁵The set $\mathcal{J}$ is dense in the set of all unimodular functions with respect to L_2 -norm. See [16] for more discussion on $\mathcal{J}$.

where g_{outer} is the outer factor of g . Since the inner-outer factorization of g is $g = g_{\text{outer}}g_{\text{inner}}$, it follows that $|g(z)/g_{\text{outer}}(z)| = |g_{\text{inner}}(z)| = 1$ for $z \in \mathbb{T}$ and therefore $q = g/\bar{g} = g/\bar{g}_{\text{outer}}$ is a unimodular function. To see $M = qH_2^-$, note that $q \in M$ hence it is enough to show $q \perp z^{-1}M$, which is equivalent to $q \perp z^{-k}g$ for all $k \geq 1$. This follows from

$$(q, z^{-k}g) = (g\bar{g}/\bar{g}_{\text{outer}}, z^{-k}) = (g_{\text{outer}}\bar{g}_{\text{outer}}/\bar{g}_{\text{outer}}, z^{-k}) \\ = (g_{\text{outer}}, z^{-k}) = 0, \quad \text{for } k \geq 1,$$

which completes the proof. $\blacksquare$

Lemma 3: A function $g \in H_2$ is non-cyclic if and only if $q \in \mathcal{J}$, where q is a unimodular function satisfying $qH_2^- = \overline{\text{span}}_{k \geq 0}\{z^{-k}g\}$.

Proof: Lemma 2 implies that the unimodular function q is on the form $q = \lambda \frac{g}{\bar{g}_{\text{outer}}}$ for some constant $|\lambda| = 1$. A function $g \in H_2$ is non-cyclic if and only if there exists a pair of inner functions φ and ψ such that

$$\frac{g}{\bar{g}} = \frac{\varphi}{\bar{\psi}} \quad \text{almost everywhere on } \mathbb{T}.$$

By combining these two facts, it follows that if $g \in H_2$ is non-cyclic, then $q = \lambda g/\bar{g}g_{\text{inner}} = \lambda \varphi/\bar{\psi}g_{\text{inner}} \in \mathcal{J}$. Conversely, if $q \in \mathcal{J}$, then

$$\frac{g}{\bar{g}} = \bar{\lambda} q g_{\text{inner}} = \frac{\varphi}{\bar{\psi}}$$

for some inner functions φ, ψ , and hence g is non-cyclic. $\blacksquare$

Lemma 4: Let q_1 and q be unimodular functions, then $q_1H_2^- \subset qH_2^-$ if and only if $\varphi q_1 = q$ for some inner function φ .

Proof: The sufficiency follows from the fact that any $g \in q_1H_2^-$ is on the form q_1f for some $f \in H_2$, and hence satisfies $g = q\bar{f}\bar{\varphi} \in qH_2^-$. To see the necessity, we note that $q_1H_2^- \subset qH_2^-$ implies $q_1 \in qH_2^-$. It follows that $q_1 = q\bar{\varphi}$ for some unimodular $\varphi \in H_2$, that is, φ is an inner function. This completes the proof. $\blacksquare$

Lemma 5: Let $q_1H_2^-$ and $q_2H_2^-$ be subspaces of L_2 where q_1 and q_2 are unimodular, then $q_1H_2^- + q_2H_2^- = L_2$ holds if and only if $q_1/q_2 \notin \mathcal{J}$.

Proof: We use proof by contradiction. Assume first that $q_1/q_2 \in \mathcal{J}$, i.e., there exist inner functions ψ_1, ψ_2 such that $q_1/q_2 = \psi_2/\psi_1$. Now, let q be the unimodular function $q = q_1\psi_1 = q_2\psi_2$. Then by Lemma 4 we have $q_jH_2^- \subset qH_2^-$ for $j = 1, 2$, and by linearity it follows that

$$q_1H_2^- + q_2H_2^- \subset qH_2^- \neq L_2.$$

Note that $qH_2^- \neq L_2$ holds since, e.g., $qz \notin qH_2^-$.

Conversely, assume that $q_1H_2^- + q_2H_2^- \neq L_2$, then $q_1H_2^- + q_2H_2^- \subsetneq L_2$ is an invariant subspace for U^{-1} while not for U (this follows since it contains an analytic function [15]). As a consequence of this there is an unimodular function q such that $q_1H_2^- + q_2H_2^- = qH_2^-$. This implies that $q_jH_2^- \subset qH_2^-$, for $j = 1, 2$. By Lemma 4 there exists inner functions ψ_1, ψ_2 such that $q = q_1\psi_1 = q_2\psi_2$, and hence $q_1/q_2 \in \mathcal{J}$. $\blacksquare$

Proof of Theorem 3: The equivalence between (a) and (b) follows directly from the Kolmogorov isomorphism. In particular, due to the Kolmogorov isomorphism, (b) implies that any linear combination of $\{w_k\}$ can be approximated by

a combination of elements in $\{x_0, x_1, \dots\}$ and $\{y_0, y_1, \dots\}$ with arbitrary accuracy; this is exactly (a). Conversely, if the system (8) is backward deterministic, then any linear combination of $\{w_k\}$ can be approximated arbitrarily by elements in $\{x_0, x_1, \dots\}$ and $\{y_0, y_1, \dots\}$, and therefore (b) holds. We next show the equivalence between (b) and (c). Using Lemma 2, it follows that (b) is equivalent to

$$q_1H_2^- + q_2H_2^- = L_2, \quad (12)$$

where $q_1 = g/\bar{g}_{\text{outer}}$ and $q_2 = h/\bar{h}_{\text{outer}}$. By Lemma 5, Equation (12) holds if and only if $q_1/q_2 \notin \mathcal{J}$. Since $q_1/q_2 = g\bar{h}h_{\text{inner}}/(\bar{g}h\bar{g}_{\text{inner}})$, where $g_{\text{inner}}, h_{\text{inner}}$ are the inner parts of g and h respectively, the equivalence with Theorem 3 (c) follows.

ACKNOWLEDGMENT

Past discussions with Anders Lindquist, Alexandre Megretski and Sergei Treil are gratefully acknowledged.

REFERENCES

- [1] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*, vol. 321. Berkeley, CA, USA: Univ. California Press, 1958.
- [2] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes: I. The regularity condition," *Acta Math.*, vol. 98, no. 1, pp. 111–150, 1957.
- [3] A. Lindquist and G. Picci, *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification* (Series in Contemporary Mathematics), vol. 1. Heidelberg, Germany: Springer, 2015.
- [4] H. Sandberg, J.-C. Delvenne, and J. C. Doyle, "On lossless approximations, the fluctuation-dissipation theorem, and limitations of measurements," *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 293–308, Feb. 2011.
- [5] T. T. Georgiou and A. Lindquist, "On time-reversibility of linear stochastic models," in *Proc. IFAC World Congr.*, vol. 19, 2014, pp. 10403–10408.
- [6] T. T. Georgiou and M. C. Smith, "Feedback control and the arrow of time," *Int. J. Control*, vol. 83, no. 7, pp. 1325–1338, 2010.
- [7] H. Tong, *Non-Linear Time Series: A Dynamical System Approach*. Oxford, U.K.: Oxford Univ. Press, 1990.
- [8] C. Diks, J. C. Van Houwelingen, F. Takens, and J. DeGoede, "Reversibility as a criterion for discriminating time series," *Phys. Lett. A*, vol. 201, nos. 2–3, pp. 221–228, 1995.
- [9] G. Weiss, "Time-reversibility of linear stochastic processes," *J. Appl. Probab.*, vol. 12, no. 4, pp. 831–836, 1975.
- [10] C. Daw, C. Finney, and M. Kennel, "Symbolic approach for measuring temporal 'irreversibility,'" *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip.*, vol. 62, no. 2, pp. 1912–1921, 2000.
- [11] 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, pp. 37–76, 1970.
- [12] P. Bloomfield, N. P. Jewell, and E. Hayashi, "Characterizations of completely nondeterministic stochastic processes," *Pac. J. Math.*, vol. 107, no. 2, pp. 307–317, 1983.
- [13] Y. A. Rozanov, "Spectral properties of multivariate stationary processes and boundary properties of analytic matrices," *Theory Probab. Appl.*, vol. 5, no. 4, pp. 362–376, 1960.
- [14] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1986.
- [15] H. Helson, *Lectures on Invariant Subspaces*. New York, NY, USA: Academic Press, 1964.
- [16] R. G. Douglas and W. Rudin, "Approximation by inner functions," *Pac. J. Math.*, vol. 31, no. 2, pp. 313–320, 1969.
- [17] Y. Chen, J. Karlsson, and T. Georgiou, "The role of past and future in estimation and the reversibility of stochastic processes," in *Proc. Int. Symp. Math. Theory Netw. Syst.*, 2014, pp. 343–346.
- [18] K. Hoffman, *Banach Spaces of Analytic Functions*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1962.