Statistical Analysis and Spectral Estimation Techniques for One-Dimensional Chaotic Signals

Steven H. Isabelle and Gregory W. Wornell, *Member, IEEE*

Abstract—Signals arising out of nonlinear dynamics are compelling models for a wide range of both natural and man-made phenomena. In contrast to signals arising out of linear dynamics, extremely rich behavior is obtained even when we restrict our attention to one-dimensional (1-D) chaotic systems with certain smoothness constraints. An important class of such systems are the so-called Markov maps. We develop several properties of signals obtained from Markov maps and present analytical techniques for computing a broad class of their statistics in closed form. These statistics include, for example, correlations of arbitrary order and all moments of such signals. Among several results, we demonstrate that all Markov maps produce signals with rational spectra, and we can therefore view the associated signals as "chaotic ARMA processes," with "chaotic white noise" as a special case. Finally, we also demonstrate how Markov maps can be used to approximate to arbitrary accuracy the statistics any of a broad class of non-Markov chaotic maps.

I. INTRODUCTION

R ECENT developments in nonlinear dynamics and chaos theory suggest that it may be possible to develop new and powerful alternative strategies for signal modeling in a variety of applications. In turn, the development of new classes of signal models may naturally lead to new kinds of algorithms for processing such signals that explicitly take into account their special structure. In this paper, we introduce and develop properties of a class of nonlinear signal models that appear to be particularly well suited to engineering applications.

The notion of using chaotic signals as models for signal processing applications has received increasing interest over the last few years [1]. There has been, for example, work on estimating the parameters of a nonlinear signal model from data and on extracting chaotic signals from noise and other forms of distortion; see, e.g., [2]–[7]. In addition, chaotic models have been proposed for a variety of practical engineering systems including sigma-delta modulators in analog-to-digital converters [8], as models for switching power converters [9] and other switched flow systems [10], and for signal generators [11]. In this paper, we focus on characterizing, analyzing, and estimating the properties of signals rather than the time

The authors are with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

series themselves or their parameters, although these issues are obviously complementary.

In our development, we restrict our attention to the case of discrete-time signals and consider scalar-valued signals $y[n]$ with the state space description

$$x[n] = F(x[n-1]) \tag{1a}$$
$$y[n] = g(x[n]) \tag{1b}$$

where

$x[0]$ initial condition,
$F(\cdot)$ nonlinear transformation that maps vectors to vectors in M-dimensional state space,
$g(\cdot)$ nonlinear transformation that maps M-dimensional state vectors into scalar observations.

To serve the broadest possible range of signal processing applications, it is tempting to avoid further constraining the structure of the signal model (1). However, as we will show, without further restriction on the properties of $F(\cdot)$ and $g(\cdot)$, the class of signals that can be described by (1) is so large as to be unwieldy for application—regardless of the dimension of the state space. As a result, in practice it is necessary to add additional constraints on the dynamics. The signals that we examine in this paper are generated by nonlinear systems that satisfy a particular class of smoothness constraints that makes them especially amenable to analysis. We demonstrate that even with such constraints and restricting our attention to a one-dimensional (1-D) state space, a rich class of signals results.

The main focus of the paper is on developing important properties of the resulting signals and efficient techniques for analyzing them. We intentionally choose an approach that deemphasizes the traditional distinctions between deterministic and stochastic signal models. In fact, the nonlinear dynamical system framework we adopt and the associated ergodic theory lends itself naturally to viewing signals from both perspectives simultaneously and makes distinctions artificial.

This paper examines the time-average statistics of signals generated by a specific class of 1-D nonlinear systems. For this class of maps, the statistics are essentially independent of the initial condition $x[0]$ that generated the time series. Moreover, if $x[0]$ is a random variable whose probability density function is appropriately chosen, the resulting random process is ergodic, i.e, its ensemble-average statistics are equal to the time-average statistics of the individual sample paths. This equivalence leads to the signal analysis tools that are the focus of this paper.

The outline of the paper is as follows. Section II explores some fundamental issues that arise in the use of nonlinear dynamics for signal modeling and makes some preliminary observations about the limitations and potential of such models. Section III establishes some basic terminology and notation and develops the basic framework by which we will obtain the statistics of signals from 1-D nonlinear maps. Section IV introduces a class of 1-D chaotic systems called Markov maps that generate signals with properties that can be characterized in an efficient vector-matrix framework. This framework leads to efficient algorithms for computing, e.g., the amplitude distributions, power spectra, and higher order statistics of such signals. In Section V, we then demonstrate that chaotic signals from Markov maps can be synthesized to approximate the behavior of any of a much broader set of chaotic signals to arbitrary accuracy. Using this important property, we then develop efficient algorithms for estimating the statistics of this larger class of chaotic signals. Finally, Section VI contains some concluding remarks, identifying open questions and directions for further research.

II. MODELING WITH CHAOS

In this section, we discuss some basic issues involved in using chaotic systems for signal modeling and, in the process, illustrate some of the potential and limitations of this approach.

We begin by observing that in the absence of further constraints, the class of signals that can be described by (1) is exceedingly large. Since this point is often not articulated in the literature, we demonstrate that, for example, it is possible to choose $F(\cdot), g(\cdot)$, and $x[0]$ to generate *any* desired time series. An immediate implication of this statement is that simply specifying that a signal was generated deterministically does not significantly limit the possible range of time series behavior. A corollary of this statement is that it is similarly possible to choose $F(\cdot), g(\cdot)$, and a probability density for $x[0]$ to generate *any* stationary stochastic process.

A construction for the case of bounded signals is as follows. First, observe that if the dimension of the state space is infinite, then it is essentially trivial to establish that there is a system $F(\cdot)$ and observation function $g(\cdot)$ that can produce any finite, causal time series. It suffices to consider the state space $\mathcal{X}$ to be the space of right-sided sequences. We use x as the (infinite-dimensional) vector notation for such a sequence and denote its ith component by $[x]_i$. In other words, each element $x \in \mathcal{X}$ is of the form

$$x = [[x]_0 \quad [x]_1 \quad [x]_2 \quad \cdots]^T.$$

Now, suppose that $F(\cdot)$ is the left-shift operator, which maps the state space to itself, and that $g(\cdot)$ observes the zeroth component of a state vector, i.e.,

$$[F(x)]_i = [x]_{i+1}, \qquad i = 0, 1, 2, \cdots$$
$$g(x) = [x]_0.$$

Then clearly, this system can generate any bounded time series by a proper choice of initial condition. In essence, by choosing the initial condition, the entire time series is chosen.

By augmenting the previous construction, stochastic processes can, likewise, be generated. In this case, the components $[x]_i$ are viewed as random variables over an appropriate probability space. Stationary processes consist of those sequences of random variables with the shift invariance property

$$\Pr\{[x]_{i_1} \in \mathcal{B}_1, [x]_{i_2} \in \mathcal{B}_2, \cdots, [x]_{i_r} \in \mathcal{B}_r\}$$
$$= \Pr\{[x]_{i_1+k} \in \mathcal{B}_1, [x]_{i_2+k} \in \mathcal{B}_2, \cdots, [x]_{i_r+k} \in \mathcal{B}_r\}$$

for all positive integers r and k and all collections of valid events $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_r$. Moreover, by a standard result of ergodic theory, every stationary process can be generated in this manner [12]. Hence, a shift can also generate any stationary random process.

It is important to note, however, that the complexity of time series behavior that can be obtained is not a function of the dimension of the state space. Indeed, the same conclusion is reached even in 1-D state spaces. To see this, it suffices to construct in 1-D state space a system equivalent to the left-shift operator of our infinite-dimensional example. This follows immediately from a classical result of real analysis—that the set of right-sided sequences of real numbers has the same cardinality as the unit interval [0, 1] (see, e.g., [13]). As a result, the points of the unit interval can be put into one-to-one correspondence with the points of the space $\mathcal{X}$. Let $\phi: [0, 1] \rightarrow \mathcal{X}$ denote this correspondence. Since ϕ is one-to-one and, hence, invertible, each state sequence generated by $F(\cdot)$ corresponds to a scalar sequence $x[n]$ in [0, 1]. The scalar sequence $x[n]$ is, therefore, determined by the dynamics

$$x[n] = \phi^{-1}(F(\phi(x[n-1]))) \triangleq f(x[n-1])$$

where $f(\cdot)$ is a scalar function mapping [0, 1] to itself. The corresponding output time series is generated by the observation equation

$$y[n] = g(\phi(x)) \triangleq h(x).$$

Thus, the 1-D map $f(\cdot)$ along with the observation function $h(\cdot)$ generate the same collection of time series as the infinite-dimensional map $F(\cdot)$ and observation function $g(\cdot)$. A similar construction could establish this type of equivalence between any two systems operating on subsets of finite-dimensional spaces.

The preceding results show that while nonlinear state space models can generate virtually any bounded time series, such models are often impractical. Indeed, the correspondence function $\phi(\cdot)$ is highly irregular, precluding its implementation with finite-precision arithmetic. However, many practical systems naturally satisfy some type of smoothness conditions. Despite the constraints implied by these smoothness conditions, a remarkably broad class of time series can be obtained and analyzed with such constrained models, even in 1-D state space.

III. ONE-DIMENSIONAL MAPS AND INVARIANT DENSITIES

For the duration of the paper, we specifically restrict our attention to discrete-time signals $x[n]$ generated by chaotic systems with a single state variable by applying the recursion

$$x[n] = f(x[n-1]) \tag{2}$$

to some initial condition $x[0]$, where the $f(\cdot)$ is a nonlinear transformation that maps scalars to scalars. Such models have been proposed not only for a variety of physical phenomena but also for engineering systems ranging from nonlinear oscillators [14] to power converters [9]. More generally, a time-series $x[n]$ can be modeled by a 1-D map whenever a scatter plot of $x[n]$ versus $x[n-1]$ resembles the graph of an appropriate function $f(\cdot)$. Such a time series can be viewed as being effectively generated by the recursion (2). Because the focus of this paper is on the analysis of signals generated by nonlinear maps, the problem of choosing an appropriate model $f(\cdot)$ will not be addressed. However, when a model exists, many applications could benefit from a detailed knowledge of the time-average properties—such as power spectra and higher order cumulants—of such a time series. While these statistics can be estimated via empirical time averaging, this approach is often computationally expensive and fails to reveal a number of special properties associated with the statistics of this class of signals.

In this section, we develop a framework for determining the time-average statistics of signals generated by 1-D maps. In particular, we will show how the time-average properties of such deterministic signals can be equated with the ensemble-average properties of the class of stationary stochastic processes generated according to the dynamics (2) with an initial condition $x[0]$ chosen from an appropriate probability distribution.

We begin by considering the properties of stochastic processes obtained from 1-D nonlinear dynamics. Let $p_0(\cdot)$ denote the probability density function of the initial condition $x[0]$, and more generally, let $p_n(\cdot)$ denote the corresponding density of the nth iterate $x[n]$. Provided that the map $f(\cdot)$ is reasonably well behaved,[1] a linear operator $P_f\{\cdot\}$ may be defined such that

$$p_n(\cdot) = P_f\{p_{n-1}(\cdot)\}.$$

This operator, which is referred to as the Frobenius–Perron (FP) operator [15], describes the time evolution of the density for the particular map.

Although, in general, the densities at distinct times n will differ, there can exist certain choices of $p_0(\cdot)$ such that the density of subsequent iterates does not change, i.e.,

$$p_0(\cdot) = p_1(\cdot) = \cdots = p_n(\cdot) \triangleq p(\cdot).$$

Such a density $p(\cdot)$, which is referred to as an *invariant density* of the map $f(\cdot)$, is a fixed point of the FP operator, i.e.,

$$p(\cdot) = P_f\{p(\cdot)\}. \tag{3}$$

For a given map $f(\cdot)$, more than one density may satisfy (3).

The invariant density plays an important role in the computation of time-averaged statistics of time series from nonlinear dynamics. When $p_0(\cdot)$ is chosen to be an invariant density, it is straightforward to verify that the resulting stochastic process is stationary and—subject to certain constraints on

---

[1] In particular, $f(\cdot)$ must be nonsingular. For the piecewise smooth maps that will be the focus of this paper, nonsingularity translates into the requirement that the derivative be nonzero almost everywhere.

$f(\cdot)$—ergodic. In this case, $f(\cdot)$ has a unique invariant density, and Birkhoff's ergodic theorem [16] can be conveniently used to establish that time averages are equivalent to ensemble averages for almost all sample waveforms. For reasons that will be discussed in Section V, we restrict our attention to this case.

The statistics of interest in this paper are correlations of the form

$$R_{f;h_0,h_1,\cdots,h_r}[k_1,\cdots,k_r]$$
$$= \lim_{L\to\infty} \frac{1}{L} \sum_{n=0}^{L-1} h_0(x[n])h_1(x[n+k_1])\cdots$$
$$h_r(x[n+k_r]) \tag{4}$$

where $x[n]$ is a time series generated by (2), the $h_i(\cdot)$ are suitably well behaved but otherwise arbitrary functions, and the k_i are nonnegative integers. This class of statistics is sufficiently broad to include as special cases the autocorrelation and all higher order moments of the time series, which are of interest in many applications.

To facilitate its computation, the correlation statistic (4) can be expressed as an ensemble average directly in terms of the invariant density and FP operator associated with the map. To see this, we begin by adopting the notation $f^n(\cdot)$ and $P_f^n\{\cdot\}$ for the respective n-fold compositions of $f(\cdot)$ and $P_f\{\cdot\}$ with themselves. Then, we rewrite (4) as

$$R_{f;h_0,h_1,\cdots,h_r}[k_1,\cdots,k_r]$$
$$= \lim_{L\to\infty} \frac{1}{L} \sum_{n=0}^{L-1} \prod_{i=0}^{r} h_i(f^{k_i}(x[n]))$$
$$= \int p(x) \prod_{i=0}^{r} h_i(f^{k_i}(x)) \, dx$$
$$= \int h_r(x)P_f^{k_r-k_{r-1}}\{h_{r-1}(x)\cdots P_f^{k_2-k_1}\{h_1(x)P_f^{k_1}$$
$$\cdot \{h_0(x)p(x)\}\}\cdots\} \, dx \tag{5}$$

where the equalities follow from applying, in turn, (2), Birkhoff's ergodic theorem, and repeated application of the identity [15]

$$\int \alpha(x)P_f\{\beta(x)\} \, dx = \int \alpha(f(x))\beta(x) \, dx$$

valid for bounded $\alpha(\cdot)$ and integrable $\beta(\cdot)$.[2] This expression for the correlation statistic will prove especially useful in computing the statistics of time series from the particular class of maps referred to as Markov maps, as we explore next.

---

[2] Note that although the FP operator is defined only for densities, its definition is extended to all integrable functions in a straightforward manner. In particular, since any integrable function $h(x)$ can be expressed in the form

$$h(x) = a_+h_+(x) - a_-h_-(x)$$

where $h_+(x)$ and $h_-(x)$ are densities and a_+ and a_- are finite, positive constants, we use

$$P_f\{h(x)\} \triangleq a_+P_f\{h_+(x)\} - a_-P_f\{h_-(x)\}.$$

IV. CHAOTIC SIGNALS FROM MARKOV MAPS

A rich class of 1-D chaotic systems that are particularly amenable to analysis are the eventually expanding, piecewise-linear, Markov maps. An adequate definition for our purposes follows.[3]

Definition 1: A map $f: [0, 1] \rightarrow [0, 1]$ is an eventually expanding, piecewise-linear, Markov map when we have the following:

1) The map is piecewise-linear, i.e., there is a set of partition points $a_0, a_1, \cdots, a_N$ satisfying $0 = a_0 < a_1 < \cdots < a_N = 1$ and such that restricted to each of the intervals $\mathcal{V}_i = (a_{i-1}, a_i)$, the map f is affine.

2) The map has the Markov property that partition points map to partition points: For each i, $f(a_i) = a_j$ for some j.

3) The map has the eventually expanding property, i.e., there is an integer $k > 0$ such that

$$\inf_{x \in [0,1]} \left| \frac{d}{dx} f^k(x) \right| > 1.$$

Despite what might appear to be a rather restrictive definition, we will see in Section V that such piecewise-linear, eventually expanding Markov maps can approximate to arbitrary accuracy any member of a much larger class of maps.

Before proceeding, we remark that the maps corresponding to Definition 1 constitute a subset of a broader class of Markov maps. These more general Markov maps also map partition points to partition points but need not be piecewise-linear nor eventually expanding [17]. Because we will not require a broader definition of Markov map in this paper, we will refer to maps satisfying Definition 1 as simply "Markov maps" when there is no risk of ambiguity.

Some additional notation and terminology will be useful in the following analysis. The intervals $\mathcal{V}_i = (a_{i-1}, a_i)$ are called partition elements. Since partition points map to partition points and $f(\cdot)$ is piecewise affine, the image of any partition element is a union of partition elements. We denote by $\mathcal{I}_i$ the set of indices of partition elements in the image of $\mathcal{V}_i$. With this notation, the image may be expressed in the form

$$f(\mathcal{V}_i) = \bigcup_{j \in \mathcal{I}_i} \mathcal{V}_j. \tag{6}$$

We will also find it useful to define the indicator function

$$\tilde{\chi}_{\mathcal{A}}(x) = \begin{cases} 1 & x \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$$

and the modified indicator function

$$\chi_i(x) = \begin{cases} 1 & x \in \mathcal{V}_i \\ 0 & \text{otherwise.} \end{cases}$$

Note that these two variants are related by

$$\chi_i(x) = \tilde{\chi}_{\mathcal{V}_i}(x).$$

---

[3] Because a map over any finite interval is equivalent to within an affine change of variables and to a map of the unit interval, we restrict our attention to maps of the unit interval with no loss of generality.
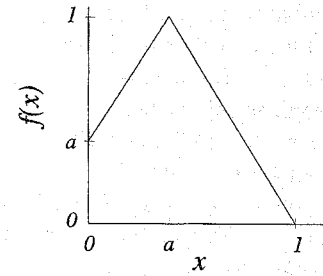

Fig. 1. Example of a piecewise-linear Markov map with two partition elements.

With these definitions, we then obtain from (6) the following identity:

$$\tilde{\chi}_{f(\mathcal{V}_i)}(x) = \sum_{j \in \mathcal{I}_i} \chi_j(x). \tag{7}$$

As an example of the use of this notation, consider the map

$$f(x) = \begin{cases} a + (1 - a)x/a & 0 \leq x \leq a \\ (1 - x)/(1 - a) & a < x \leq 1 \end{cases} \tag{8}$$

which meets the conditions of Definition 1 and is depicted in Fig. 1. This map has partition elements $\mathcal{V}_1 = [0, a]$ and $\mathcal{V}_2 = [a, 1]$. The index sets associated with the partition elements are $\mathcal{I}_1 = \{2\}$ and $\mathcal{I}_2 = \{1, 2\}$. We will make use of this example on a number of occasions throughout the paper.

Markov maps have a number of important properties that make them attractive for applications. For example, all Markov maps have invariant densities and are ergodic under readily verifiable conditions [17]. In addition, suitably quantized outputs of Markov maps are equivalent to Markov chains. In particular, for almost all initial conditions, the sequence of partition element indices corresponding to successive iterates of the map is indistinguishable from a sample path of a Markov chain [18].

A. Statistics of Markov Maps

In this section, we establish another property of Markov maps—that their statistics can be determined in closed form. To develop a strategy for computing these statistics, we begin by noting that from Definition 1, we may express a Markov map in the form

$$f(x) = \sum_{i=1}^{N} (s_i x + b_i) \chi_i(x) \tag{9}$$

where $s_i \neq 0$ for all i. This leads to a convenient representation for the map's FP operator, as we now develop.

1) Frobenius–Perron Operator: It is straightforward to establish that the FP operator for nonsingular, 1-D maps of the form

$$f(x) = \sum_{i=1}^{N} f_i(x) \chi_i(x)$$

where each $f_i(\cdot)$ is continuous and monotonic, can be expressed in the form [15]

$$P_f\{h(x)\} = \sum_{i=1}^{N} \frac{h(f_i^{-1}(x))\tilde{\chi}_{f_i(\mathcal{V}_i)}(x)}{|f'(f_i^{-1}(x))|}. \qquad (10)$$

The FP operator for a piecewise linear map then follows by substituting (9) into (10) to obtain

$$P_f\{h(x)\} = \sum_{i=1}^{N} \frac{1}{|s_i|} h\left(\frac{x-b_i}{s_i}\right)\chi_{f(\mathcal{V}_i)}(x). \qquad (11)$$

One unique property of FP operators of piecewise-linear Markov maps is that their invariant subspaces contain piecewise polynomials of the form

$$h(x) = \sum_{i=1}^{N}\sum_{j=0}^{K} a_{ij}x^j \chi_i(x) \qquad (12)$$

where the a_{ij} are arbitrary scalars. To see this, note that such functions can be expanded in the basis

$$\{\theta_1(x), \theta_2(x), \cdots, \theta_{N(K+1)}\}$$
$$\triangleq \{\chi_1(x), \cdots, \chi_N(x), x\chi_1(x), \cdots, x\chi_N(x), \cdots,$$
$$x^K\chi_1(x), \cdots, x^K\chi_N(x)\} \qquad (13)$$

to yield the representation

$$h(x) = \sum_{i=1}^{N(K+1)} h_i\theta_i(x).$$

Thus, each such function can be uniquely represented by the vector

$$h = [h_1 \quad h_2 \quad \cdot \quad h_{N(K+1)}]^T$$

which we refer to as the coordinate vector of $h(x)$. We denote by $\mathcal{P}_K$ the $N(K+1)$-dimensional space spanned by the basis functions $\{\theta_i\}_{i=1}^{N(K+1)}$.

For these piecewise polynomial functions, a convenient expression for $P_f\{h(x)\}$ can be developed by using (11) to obtain

$$P_f\{\theta_{jN+i}(x)\} = P_f\{x^j\chi_i(x)\}$$
$$= \sum_{l=1}^{N} \frac{1}{|s_l|}\left(\frac{x-b_l}{s_l}\right)^j \chi_i\left(\frac{x-b_l}{s_l}\right)\chi_{f_l(I_l)}(x). \qquad (14)$$

After exploiting the relation (7) and the Markov property of the map $f(\cdot)$, (14) simplifies to

$$P_f\{\theta_{jN+i}(x)\} = \left(\frac{x-b_i}{s_i}\right)^j \frac{1}{|s_i|}\sum_{l\in\mathcal{I}_i}\chi_l(x). \qquad (15)$$

Finally, using (15) and exploiting the linearity of the FP operator, we obtain, for $h(\cdot)$ of the form (12), that

$$P_f\{h(x)\} = \sum_{i=1}^{N}\sum_{j=0}^{K} a_{ij}\left(\frac{x-b_i}{s_i}\right)^j \frac{1}{|s_i|}\sum_{l\in\mathcal{I}_i}\chi_l(x). \qquad (16)$$

From (16), we see that the images of piecewise-polynomial functions $h(\cdot)$ of degree K are also piecewise polynomials of degree K, i.e., the operator $P_f\{\cdot\}$ maps $\mathcal{P}_K$ to itself. Since $\mathcal{P}_K$ is a finite-dimensional space, the restriction of $P_f\{\cdot\}$ to $\mathcal{P}_K$ can be represented by a square $N(K+1)$-dimensional matrix, which we will denote by $\boldsymbol{P}_K$. This matrix, which we will refer to as the FP matrix, describes how the coefficients of expansions in terms of the basis (13) map under the FP operator. In particular, if

$$P_f\{h(x)\} \triangleq q(x) = \sum_{j=1}^{N(K+1)} q_j\theta_j(x)$$

then

$$q = P_K h \qquad (17)$$

where

$$q = [q_1 \quad q_2 \quad \cdot \quad q_{N(K+1)}]^T.$$

Using (16) with the binomial theorem, it follows immediately that $\boldsymbol{P}_K$ is block upper triangular. In particular

$$P_K = \begin{bmatrix} P_{00} & P_{01} & \cdots & \cdots & P_{0K} \\ 0 & P_{11} & P_{12} & \cdots & P_{1K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & P_{KK} \end{bmatrix}$$

where each nonzero $N \times N$ block is of the form

$$P_{ij} = \binom{j}{i} P_0 B^{j-i}S^j \quad \text{for} \quad j \geq i.$$

The $N \times N$ matrices B and S are diagonal with elements $B_{ii} = -b_i$ and $S_{ii} = 1/s_i$, respectively, while $P_0 = P_{00}$ is the $N \times N$ matrix with elements

$$[P_0]_{ij} = \begin{cases} 1/|s_j| & i\in\mathcal{I}_j \\ 0 & \text{otherwise.} \end{cases}$$

The matrix P_0 has some key properties that will be important in our subsequent development. In particular, as established by Friedman and Boyarsky [19], P_0 is diagonally similar to a column stochastic matrix, i.e., there is a diagonal matrix D with positive entries such that $D^{-1}P_0D$ has positive elements, and each of its columns sums to unity.

Since similar matrices have the same eigenvalues, it follows that P_0 has the same eigenvalues as a stochastic matrix. However, via Frobenius's theorem [20], all stochastic matrices—and, hence, P_0—have eigenvalues with magnitude not greater than unity and at least one eigenvalue equal to unity. In turn, we can conclude that

$$\|P_0\| \leq 1$$

where $\|\cdot\|$ denotes the usual matrix norm [21]. Furthermore, because the elements of P_0 are nonnegative, an eigenvector corresponding to a unit eigenvalue has nonnegative components.

2) Invariant Densities: One use of the matrix representation of the FP operator is for directly calculating the invariant densities of a Markov map, which is a problem first considered in [22] and later in [17]. Invariant densities are obtained by solving a matrix eigenvalue problem, as we now develop.

A nonnegative, unit-area function $h(x)$ is an invariant density if its coordinate vector is a solution to (17) with $q(x) = h(x)$. Markov maps have the property that all invariant densities of interest are vectors in $\mathcal{P}_0$ [23], i.e., they are piecewise constant and can be expressed in the form

$$p(x) = \sum_{j=1}^{N} p_j \theta_j(x). \tag{18}$$

It follows that the coordinate vector of the invariant density

$$\boldsymbol{p} = [p_1 \quad p_2 \quad \cdots \quad p_N]^T$$

is obtained as the solution to the eigenvector problem

$$\boldsymbol{P}_0 \boldsymbol{p} = \boldsymbol{p}.$$

Thus, this coordinate vector is the nonnegative eigenvector of $\boldsymbol{P}_0$ corresponding to the unit eigenvalue discussed in Section IV-A1.

3) Correlation Statistics: Let us next consider the computation of more general correlation statistics (4) when each $h_i(\cdot)$ is a piecewise polynomial, i.e., $h_i \in \mathcal{P}_K$ for each i. First, we express (4) using (5) as

$$R_{f;h_0,h_1,\cdots,h_r}[k_1,\cdots,k_r] = \int g_1(x) g_2(x)\, dx \tag{19}$$

where

$$g_1(x) = h_r(x) \tag{20a}$$

$$g_2(x) = P_f^{k_r - k_{r-1}} \{ h_{r-1}(x) \cdots P_f^{k_2 - k_1} \cdot \{ h_1(x) P_f^{k_1} \{ h_0(x) p(x) \} \} \cdots \}. \tag{20b}$$

Then, expanding $g_1(x)$ and $g_2(x)$ in the basis (13), i.e.,

$$g_1(x) = \sum g_{1,i} \theta_i(x) \tag{21a}$$

$$g_2(x) = \sum g_{2,j} \theta_j(x) \tag{21b}$$

the associated coordinate vectors can be expressed in the form

$$[g_{1,1} \quad g_{1,2} \quad \cdots \quad g_{1,N(K+1)}]^T \triangleq \boldsymbol{g}_1 = \boldsymbol{h}_r \tag{22a}$$

and

$$[g_{2,1} \quad g_{2,2} \quad \cdots \quad g_{2,N(K+1)}]^T \triangleq \boldsymbol{g}_2 \\ = \boldsymbol{P}^{k_r - k_{r-1}}(\boldsymbol{h}_{r-1} \odot \cdots \odot \boldsymbol{P}^{k_2 - k_1}(\boldsymbol{h}_2 \odot \boldsymbol{P}^{k_1} \\ \cdot (\boldsymbol{h}_1 \odot \boldsymbol{P})) \cdots) \tag{22b}$$

where $\boldsymbol{P}$ is the FP matrix of sufficient dimension, and where $\odot$ denotes the polynomial product operator. For two coordinate vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, the notation $\boldsymbol{u}_1 \odot \boldsymbol{u}_2$ denotes the coordinate vector for the corresponding product of piecewise polynomials $u_1(x) u_2(x)$ in a basis of suitably high dimension.

In turn, using (21) in (19) with a basis of sufficient dimension, we obtain

$$R_{f;h_0,h_1,\cdots,h_r}[k_1,\cdots,k_r] = \boldsymbol{g}_1^T \boldsymbol{M} \boldsymbol{g}_2 \tag{23}$$

where $\boldsymbol{M}$ is a basis correlation matrix with elements

$$[\boldsymbol{M}]_{ij} = \int \theta_i(x) \theta_j(x)\, dx.$$

When it is important to make the dimension of this matrix explicit, we will use the notation $\boldsymbol{M}_K$ when $\boldsymbol{M}$ is $N(K+1) \times N(K+1)$, corresponding to the case of Kth-order piecewise polynomials. Note that each row and column of $\boldsymbol{M}_K$ contains at most $K + 1$ nonzero entries because the support of only $K+1$ of the θ_i's coincide, and the rest are disjoint. This sparse structure of $\boldsymbol{M}$ can be exploited to obtain computationally efficient implementations of (23).

4) Power Spectra: We now apply the results of the preceding section to problem of computing the power spectral densities of time series generated by Markov maps. This yields an algorithm for computing such spectra, as well as insight into the properties such signals. We focus on the special case of power spectra both because of its familiarity and because power spectra have a strong physical significance in many engineering systems. As an important example of the latter, the power spectra of chaotic currents in power converters governed by nonlinear dynamics is directly related to their voltage ripple and to the spectrum of emitted radiation (see, e.g., [6] and the references therein).

To begin, we express the autocorrelation function for the time series in terms of the FP operator using the general expression (5) to obtain

$$R_{xx}[k] = E\{x[n]x[n+k]\} = E\{x f^k(x)\} \\ = \int x f^k(x) p(x)\, dx = \int x P_f^k \{ x p(x) \}\, dx$$

where $p(\cdot)$ is the invariant density associated with $f(\cdot)$. Since $R_{xx}[k]$ is symmetric, we restrict our attention to $k \geq 0$ and obtain that $g_1(x)$ and $g_2(x)$ in (20) specialize to

$$g_1(x) = x \tag{24a}$$

$$g_2(x) = P_f^k \{ x p(x) \}. \tag{24b}$$

To determine the dimension required in the basis expansion (21), recall that the invariant density of a Markov map is piecewise constant (see Section IV-A2), and hence, product $x p(x)$ is piecewise linear. Since $P_f\{\cdot\}$ maps piecewise-linear functions to piecewise-linear functions as developed in Section IV-A1, we have that $g_2(x)$ is piecewise linear. Combining this with the fact that $g_1(x) = x$ is, of course, linear, we see that the space of piecewise-linear polynomials $\mathcal{P}_1$ suffices for the expansion. Accordingly, the FP matrix $\boldsymbol{P}_1$ and basis correlation matrix $\boldsymbol{M}_1$ can be used, both of which are of size $2N \times 2N$.

Using (23) and (22), the correlation sequence can be expressed in the form

$$R_{xx}[k] = \boldsymbol{g}_1^T \boldsymbol{M}_1 \boldsymbol{P}_1^k \tilde{\boldsymbol{g}}_2 \tag{25}$$

where each $\boldsymbol{g}_1$ is the coordinate vector associated with (24) and where $\tilde{\boldsymbol{g}}_2$ is the coordinate vector associated with

$$\tilde{g}_2(x) = x p(x)$$

since

$$g_2(x) = P_f^k(\tilde{g}_2(x)).$$

To obtain the power spectrum, we take the Fourier transform of (25), which yields, exploiting symmetry

$$S_{xx}(e^{j\omega}) = g_1^T M_1 \left(\sum_{k=-\infty}^{+\infty} P_1^{|k|} e^{-j\omega k} \right) \tilde{g}_2. \qquad (26)$$

Further simplification of (26) is obtained by examining the structure of the FP matrix P_1. In general, P_1 has eigenvalues whose magnitude is strictly less than unity and others with unit magnitude and at least one unit eigenvalue—that associated with the invariant density as discussed in Section IV-A2. The unit-magnitude eigenvalues of P_1 give rise to impulses in the Fourier transform (26). The unit eigenvalue gives rise to the DC component or, equivalently, the mean of the time series.

The effects of eigenvalues of unit magnitude on the spectrum can be isolated by expressing P_1 in the Jordan form

$$P_1 = E^{-1}JE$$

where E is a matrix of generalized eigenvectors and where

$$J = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}$$

with J_1 consisting of Jordan blocks with eigenvalues of unit magnitude and J_2 consisting of Jordan blocks with the remaining eigenvalues, whose magnitudes are strictly less than unity. In turn, P_1 can be expressed in the form

$$P_1 = \Gamma_1 + \Gamma_2$$

where Γ_1 and Γ_2 are defined by

$$\Gamma_1 = E^{-1} \begin{bmatrix} J_1 & 0 \\ 0 & 0 \end{bmatrix} E$$

$$\Gamma_2 = E^{-1} \begin{bmatrix} 0 & 0 \\ 0 & J_2 \end{bmatrix} E.$$

It can be readily verified that $P_1^k = \Gamma_1^k + \Gamma_2^k$ so that the sum (26) is of the form

$$S_{xx}(e^{j\omega}) = h_1^T M (I - \Gamma_2 e^{-j\omega})^{-1} (I - \Gamma_2^2)(I - \Gamma_2 e^{j\omega})^{-1} \tilde{g}_2$$
$$+ \sum_{i=1}^{m} C_i \delta(\omega - \omega_i) \qquad (27)$$

where C_i and ω_i depend on Γ_1, and m is no larger than the dimension of J_1. Assuming the process generated by $f(\cdot)$ has a nonzero mean, $m \geq 1$, and $\omega_i = 0$ for some i with $C_i \neq 0$.

From (27), we can conclude that the spectrum of a Markov map is a linear combination of an impulsive component and a rational function. This result implies that there are classes of rational spectra that can be generated not only by the usual synthesis of driving white noise through a linear time-invariant filter with a rational system function but also by iterating deterministic nonlinear dynamics. Thus, it is natural to view chaotic signals corresponding to Markov maps as "chaotic autoregressive moving-average (ARMA) processes." Note that a special case corresponds to the "chaotic white
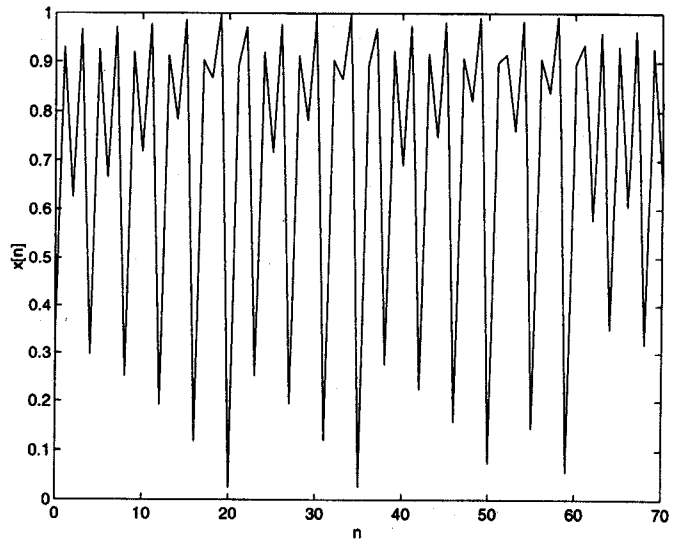

Fig. 2. Representative time series generated by the Markov map of Fig. 1 for $a = 8/9$.

noise" described in [3] and the first-order autoregressive processes described in [24].

A natural question is whether an arbitrary rational spectrum can be obtained via Markov map dynamics with a suitable map. The answer to this question remains open, although the answer is probably negative due to the complicated dependency of pole and zero locations on the map. We point out, for instance, that the poles of the rational portion of the spectrum correspond to the eigenvalues of the matrix Γ_2, i.e., the eigenvalues of P_1 with magnitude less than one. The zeros of $S_{xx}(z)$ depend on the vectors g_1 and $\tilde{g}_2$ and the matrix M.

B. An Example

In this section, we consider a simple example that illustrates the use of these techniques. In particular, we return to the Markov map defined in (8) and depicted in Fig. 1. A representative time series generated by iterating $f(\cdot)$ from the initial condition $x[0] = 1/3$ is shown Fig. 2 for the case $a = 8/9$.

When restricted to the 2-D space corresponding to piecewise-constant functions, the FP matrix associated with $f(\cdot)$ is readily calculated to be

$$P_0 = \begin{bmatrix} 0 & 1-a \\ a/(1-a) & 1-a \end{bmatrix}.$$

Furthermore, the eigenvector associated with the unique unit eigenvalue is

$$p = [1-a \quad 1]^T.$$

Via (18), we then get that the invariant density is the normalized piecewise-constant function

$$p(x) = \begin{cases} 1/(1+a) & 0 \leq x \leq a \\ 1/(1-a^2) & a \leq x \leq 1. \end{cases}$$

In Fig. 3, this density is compared with an empirically estimated density computed via a histogram of 50 000 points generated by the map f for the case $a = 8/9$.
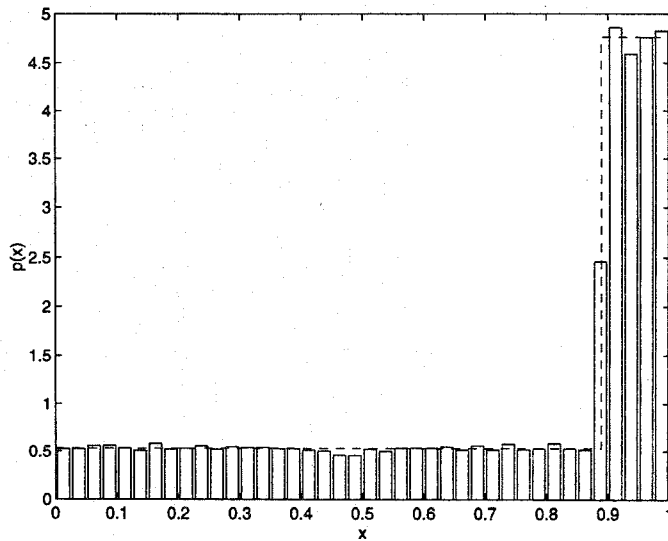
Fig. 3. Comparison of the analytically computed first-order density with an empirical estimate for the Markov map of Fig. 1. The analytically computed density is indicated via the dashed curve. The empirical estimate, obtained via a histogram with 40 bins applied to a time series of length 50 000, is indicated via the bars.
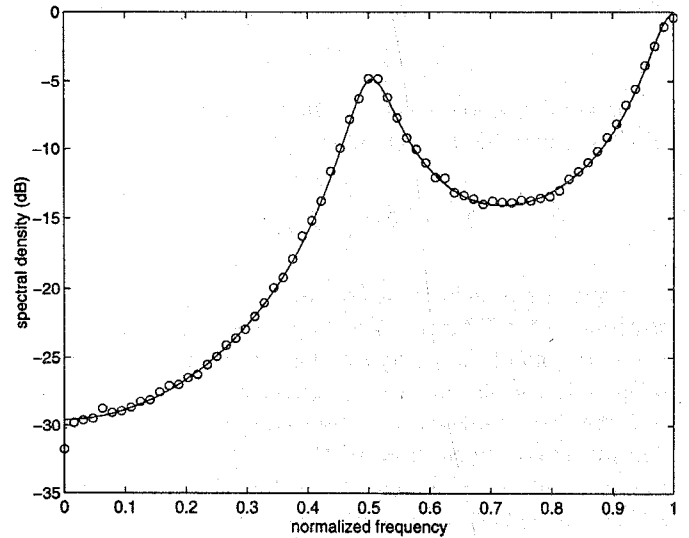

Fig. 4. Comparison of analytically computed power spectrum to empirical power spectrum for the map of Fig. 1. The solid line indicates the analytically computed spectrum, while the circles indicate the samples of the spectrum estimated by applying periodogram averaging to a time series of length 50 000.

To compute the power spectrum for this example map, we first compute the FP matrix associated with the FP operator on the space of piecewise-linear functions $\mathcal{P}_1$. From this calculation, we obtain

$$P_1 = \begin{bmatrix} 0 & 1-a & 0 & 1-a \\ a/(1-a) & 1-a & -a^3/(1-a)^2 & (1-a) \\ 0 & 0 & 0 & -(1-a)^2 \\ 0 & 0 & a^2/(1-a)^2 & -(1-a)^2 \end{bmatrix}.$$

In turn, we compute the basis correlation matrix corresponding to the same space, obtaining

$$M_1 = \begin{bmatrix} a & 0 & a^2/2 & 0 \\ 0 & 1-a & 0 & (1-a^2)/2 \\ a^2/2 & 0 & a^3/3 & 0 \\ 0 & (1-a^2)/2 & 0 & (1-a^3)/3 \end{bmatrix}.$$

Specializing our results to the case $a = 8/9$, we first obtain that the average value of the time series is $m_x = 217/306$. Then, using (27), we obtain that the rational part of the z transform of the correlation sequence is given by (28), shown at the bottom of the page.

The power spectrum corresponding to evaluating (28) on the unit circle $z = e^{j\omega}$ is plotted in Fig. 4, along with an empirical spectrum computed by periodogram averaging with a window length of 128 on a time series of length 50 000. The solid line corresponds to the analytically obtained expression (28), whereas the circles represent the spectral samples estimated by periodogram averaging.

V. MODELING WITH MARKOV MAPS

A much larger class of chaotic signals for applications are those corresponding to general *eventually expanding* maps, which are defined as follows. Without loss in generality, we again restrict our attention to maps of the unit interval.

Definition 2: A nonsingular map $f: [0, 1] \rightarrow [0, 1]$ is called eventually expanding if the following hold:

1) There is a set of partition points $0 = a_0 < a_1 < \cdots a_N = 1$ such that restricted to each of the intervals $\mathcal{V}_i = (a_{i-1}, a_i)$, the map $f(\cdot)$ is monotonic, continuous, and differentiable.
2) The function $1/|f'(x)|$ is of bounded variation [13].[4]
3) There exists a real $\lambda > 1$ and an integer m such that

$$\left| \frac{d}{dx} f^m(x) \right| \geq \lambda$$

for all x wherever the derivative exists. This is the eventually expanding condition.

The class of eventually expanding maps includes not only the piecewise-linear Markov maps of Definition 1 as a small subset but also a wide range of maps that are not piecewise-linear, do not satisfy the Markov property of mapping partition points to partition points, and/or do not have a bounded derivative. Note from the definition that an eventually expanding map need not be continuous—only piecewise so. Non-Markov

[4] This is a smoothness condition on the derivative. In some definitions, this condition is replaced with a more restrictive bounded slope condition, i.e., that there exist a constant B such that $|f'(x)| < B$ for all x. However, in our case, it is just as convenient to consider the broader class of eventually expanding maps.

$$S_{xx}(z) = -\frac{42632}{459} \frac{36z^{-1} - 145 + 36z}{(9 + 8z)(9 + 8z^{-1})(64z^2 + z + 81)(64z^{-2} + z^{-1} + 81)}. \tag{28}$$

eventually expanding maps arise in a number of applications. As one example, they provide useful models for inductor current in practical switching power converters [6] that can be substantially more accurate than traditional linear models.

Some properties of eventually expanding maps have been well-studied. For example, conditions for ergodicity and the applicability of central limit theorems have been developed [25]. It is known that all eventually expanding maps have invariant densities [26]. In particular, an eventually expanding map with N partition points has at least one and, at most, N invariant densities, all of which are of bounded variation. Moreover, as discussed in [26], certain auxiliary conditions can be imposed on $f(\cdot)$ to ensure that a unique invariant density exists. However, it is useful to note that when $f(\cdot)$ has multiple invariant densities, each is supported on disjoint subsets of the unit interval consisting of finite unions of intervals. When restricted to one of these subsets, $f(\cdot)$ is an ergodic map, i.e., it has a unique invariant density. In this manner, any eventually expanding map may be decomposed into a finite number of ergodic components. Using this decomposition, maps with multiple invariant densities can be analyzed by examining each ergodic component separately.[5]

In this section, we explore the statistics of time series generated by eventually expanding maps. These properties have received comparatively little attention in the literature. In particular, we show that the statistics of time series from eventually expanding maps can be approximated to arbitrary accuracy by those of piecewise-linear, eventually expanding Markov maps. As we will see, such approximation strategies provide a powerful method for analyzing chaotic time series from eventually expanding maps. For simplicity of exposition, we restrict our attention to ergodic maps, i.e., maps for which there is a single ergodic component.

We begin by considering a sequence of piecewise-linear Markov maps $\hat{f}_i(\cdot)$ and examining conditions under which the statistics of $\hat{f}_i(\cdot)$ converge to those of a given eventually expanding map $f(\cdot)$. This important mode of convergence, which we will refer to as *statistical convergence*, is formally defined as follows.

Definition 3: Let $f(\cdot)$ be an eventually expanding map with a unique invariant density $p(\cdot)$. A sequence of maps $\{\hat{f}_i(\cdot)\}$ statistically converges to $f(\cdot)$ if each $\hat{f}_i(\cdot)$ has a unique invariant density $p_i(\cdot)$, and

$$R_{\hat{f}_i, h_0, h_1, \cdots, h_r}[k_1, \cdots, k_r]$$
$$\rightarrow R_{f, h_0, h_1, \cdots, h_r}[k_1, \cdots, k_r] \quad \text{as } i \rightarrow \infty$$

for any continuous $h_j(\cdot)$ and all finite k_j and finite r.

A sequence of Markov maps that statistically converges to a given eventually expanding map $f(\cdot)$ can be constructed in a computationally straightforward manner. To begin, we denote by $\mathcal{Q}$ the set of partition points of $f(\cdot)$ and by $\mathcal{Q}_i$ the set of partition points of the ith in the sequence of Markov map approximations. The sets of partition points for the increasingly fine approximations are defined recursively via

$$\mathcal{Q}_i = \mathcal{Q}_{i-1} \cup f^{-1}(\mathcal{Q}_{i-1}).$$

[5] See, e.g., [25] for a more detailed treatment of these issues.

In turn, each $\hat{f}_i(\cdot)$ is defined by specifying its value at the partition points $\mathcal{Q}_i$ according to the following prescription:[6]

1) For each partition point $q \in \mathcal{Q}_i$ such that $f(\cdot)$ is increasing at q^+, define $\hat{f}_i(q^+)$ by

$$\hat{f}_i(q^+) = \max \{v \in \mathcal{Q}_i | v \leq f(q^+)\}.$$

2) For each partition point $q \in \mathcal{Q}_i$ such that $f(\cdot)$ is decreasing at q^+, define $\hat{f}_i(q^+)$ by

$$\hat{f}_i(q^+) = \min \{v \in \mathcal{Q}_i | v \geq f(q^+)\}.$$

3) For each partition point $q \in \mathcal{Q}_i$ such that $f(\cdot)$ is increasing at q^-, define $\hat{f}_i(q^-)$ by

$$\hat{f}_i(q^-) = \min \{v \in \mathcal{Q}_i | v \geq f(q^-)\}.$$

4) For each partition point $q \in \mathcal{Q}_i$ such that $f(\cdot)$ is decreasing at q^-, define $\hat{f}_i(q^-)$ by

$$\hat{f}_i(q^-) = \max \{v \in \mathcal{Q}_i | v \leq f(q^-)\}.$$

At all other points, the map $\hat{f}_i(\cdot)$ is defined by linear interpolation. Note that the number of partition points generally grows exponentially as i increases. We emphasize that with this construction, each approximating function $\hat{f}_i(\cdot)$ is not only piecewise linear but also satisfies the Markov property, i.e., $\hat{f}_i(\mathcal{Q}_i) \subset \mathcal{Q}_i$ and the eventually expanding property [27]. Hence, each is a Markov map in the sense of Definition 1.

For this sequence of approximations, we have the following theorem, whose proof is provided in Appendix A.

Theorem 1: Suppose $f(\cdot)$ is an eventually expanding map in the sense of Definition 2. Then, the sequence of piecewise-linear Markov approximations $\hat{f}_i(\cdot)$ statistically converges to $f(\cdot)$, i.e., converges in the sense of Definition 3.

We remark that this theorem includes as a special case the results of [28], in which such approximations were used to estimate the Lyapunov exponent of a 1-D map.

From Theorem 1, we conclude that for sufficiently large i, the statistics of $\hat{f}_i(\cdot)$ are close to those of $f(\cdot)$. The practical consequence of this result is that correlation statistics of an eventually expanding map $f(\cdot)$ can be approximated by first determining a Markov map $\hat{f}_k(\cdot)$ that is a good approximation to $f(\cdot)$ and then finding the statistics of Markov map using the techniques described in Section IV-A3.

This approach can be used, in particular, to efficiently estimate the power spectrum associated with an eventually expanding map. As an example, consider the map shown in Fig. 5, which is not piecewise-linear and does not satisfy the Markov property of Definition 1. As illustrated in Fig. 6, a time-series generated by this map alternates irregularly between periods of exponential growth in amplitude and periods of rapid decay, which results in a pronounced peak in its power spectrum. Fig. 7 shows an empirically computed power spectrum along with two approximate spectra. The dashed and solid lines correspond to approximate spectra computed analytically using the Markov maps $\hat{f}_1(\cdot)$ and $\hat{f}_3(\cdot)$, respectively, whereas the dash-dotted line corresponds to an

[6] Note that since the original map $f(\cdot)$ can be discontinuous, we use superscripts $+$ and $-$ to denote the left and right limiting values of the map at an individual partition point.
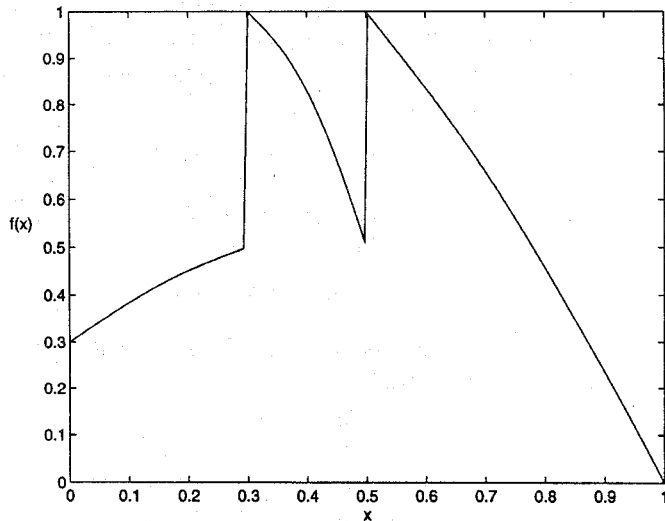
Fig. 5. Example of an eventually expanding map that is neither piecewise linear nor Markov.
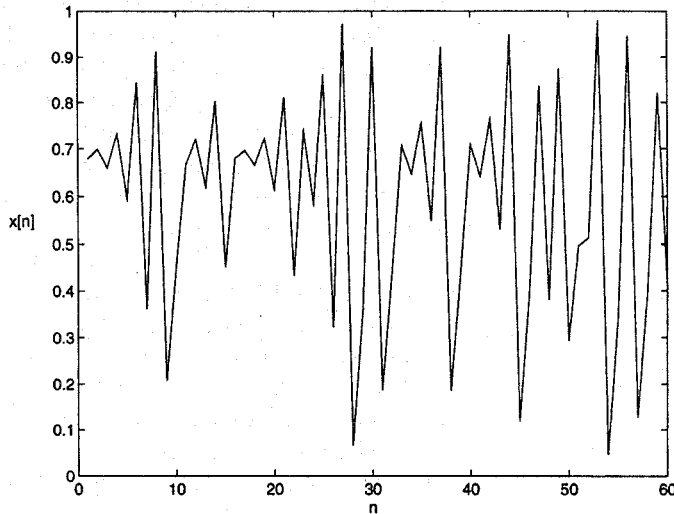

Fig. 6. Segment of a typical time-series generated by the map in Fig. 5.

estimate of the power spectrum associated with $f(\cdot)$ obtained by applying periodogram averaging to a time-series generated by $f(\cdot)$.

Not surprisingly, computing spectral estimates via Markov map approximations requires, in general, significantly less computation than those obtained via periodogram averaging. Furthermore, this technique avoids numerical difficulties inherent in simulating chaotic systems on digital computers.

VI. CONCLUDING REMARKS

In this paper, new classes of signal models have been developed and analyzed. The properties of such models suggest that they may be efficient and convenient alternatives to many conventional ARMA process models that are so widely used in signal processing applications. In particular, we showed that chaotic signals generated by deterministic nonlinear dynamics corresponding to piecewise-linear, eventually expanding Markov maps share important properties
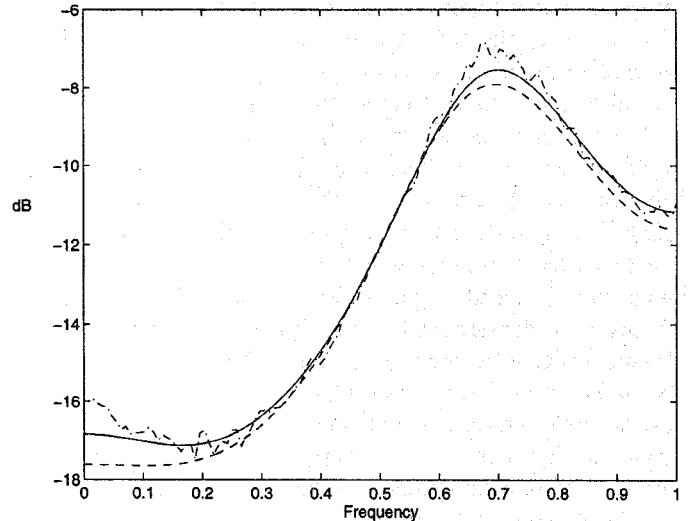

Fig. 7. Comparison of approximate and empirical power spectra for the map in Fig. 5. The dashed line and solid line correspond to approximate spectra computed analytically using the Markov maps $\hat{f}_1(\cdot)$ and $\hat{f}_3(\cdot)$, respectively, whereas the dash-dotted line corresponds to an estimate of the power spectrum associated with $f(\cdot)$ obtained by applying periodogram averaging to a time-series generated by $f(\cdot)$.

with ARMA processes. In particular, all such signals have rational spectra. For these reasons, it is convenient to view these chaotic signals as "chaotic ARMA processes." While we explored several properties of these processes, many others remain to be investigated. Moreover, many interesting and important questions remain open. For example, as discussed in Section IV-A4, while all chaotic ARMA processes have rational spectra, it is unclear whether a chaotic ARMA process can be constructed to produce any desired spectrum. Such a result would, naturally, have important implications. Likewise, exploring the properties of higher order spectra for these signals may lead to important insights and applications.

We also established results to show that chaotic ARMA processes can be constructed to approximate the statistical behavior of a broad class of chaotic processes to arbitrary accuracy. In particular, we constructed a sequence of piecewise-linear Markov maps whose statistics all converge to those of a desired arbitrary eventually expanding map. These results are particularly useful in applications where such maps arise naturally, such as in the case of switching power converters [6]. In such applications, the approximations allow important statistics of these chaotic signals that cannot be computed in closed form to be efficiently estimated; an important example is power spectrum estimation. Important avenues for further research remain. For example, it would be worthwhile to explore whether other useful sequences of Markov map approximations can be obtained for modeling eventually expanding maps. Of particular interest would be alternative sequences that yield better approximations with fewer partition points since this would reduce modeling complexity. Whether such approximating sequences exist is an open question. A related issue is that of characterizing the error inherent in our approximation procedure. In this area, bounds on the approximation error expressed in terms of the approximating sequence would be desirable.

APPENDIX A
PROOF OF THEOREM 1

To prove our result, we need to exploit the notion of weak convergence, which is defined as follows [15].

Definition 4: A sequence of functions $\phi_i \in L_1$ converges weakly to $\phi \in L_1$ if

$$\lim_{i \to \infty} \int_{[0,1]} h(x)\phi_i(x)\,dx = \int_{[0,1]} h(x)\phi(x)\,dx \qquad (29)$$

for all bounded measurable functions h.

In addition, we will exploit the following theorem due to Gora [27].

Theorem 2: Suppose f is an eventually expanding map with invariant density p, and $\hat{f}_i$ is the sequence of Markov approximation to f. If $\hat{f}_i$ has invariant density $\hat{p}_i$, then $\hat{f}_i \to f$ uniformly and $\hat{p}_i \to p$ weakly.

To obtain our main result, we begin by expressing the difference between the true and approximating statistics in the form

$$
\begin{aligned}
R_{\hat{f}_i, h_0, h_1, \cdots, h_r} & [k_1, \cdots, k_r] - R_{f, h_0, h_1, \cdots, h_r}[k_1, \cdots, k_r] \\
&= \int h_0(x)h_1(\hat{f}_i^{k_1}(x)) \cdots h_r(\hat{f}_i^{k_r}(x))\hat{p}_i(x)\,dx \\
&\quad - \int h_0(x)h_1(f^{k_1}(x)) \cdots h_r(f^{k_r}(x))p(x)\,dx \\
&= \int h_0(x)h_1(\hat{f}_i^{k_1}(x)) \cdots h_r(f^{k_r}(x))[\hat{p}_i(x) - p(x)]\,dx \\
&\quad + \int [h_0(x)h_1(\hat{f}_i^{k_1}(x)) \cdots h_r(\hat{f}_i^{k_r}(x)) - h_0(x) \\
&\quad \cdot h_1(f^{k_1}(x)) \cdots h_r(f^{k_r}(x))]\hat{p}_i(x)\,dx.
\end{aligned} \qquad (30)
$$

The first term in the second equality in (30) goes to zero as $i \to \infty$ by the weak convergence of $\hat{p}_i(\cdot)$ to $p(\cdot)$ that is guaranteed by Theorem 2. The second term goes to zero as $i \to \infty$, provided two conditions are met: 1) Each $h_j(\cdot)$ is uniformly continuous, and 2) each $\hat{f}_i^{k_j}$ converges uniformly to $f^{k_j}(\cdot)$.

The first of these conditions follows from the fact since each h_j is continuous on [0, 1], it is also uniformly continuous. The second of these conditions is established by the following lemma, which completes our proof.

Lemma 1: The sequence of Markov approximations has the property that $\hat{f}_i^j$ converges uniformly to f^j as $i \to \infty$ for each nonnegative j.

Proof: When f is continuous, the uniform convergence of $\hat{f}_i$ to f immediately implies that $\hat{f}_i^j$ converges uniformly to f^j. In the remainder of the proof, we therefore consider the case in which f is not continuous.

We denote the set of initial partition points by $\mathcal{Q}_0 = \{a_j\}_{j=1}^N$ and the set corresponding to the ith in the sequence of approximating maps by $\mathcal{Q}_i = \mathcal{Q}_{i-1} \cup f^{-1}(\mathcal{Q}_{i-1}) = \{a_j^{(i)}\}$. Moreover, we denote the associated partitions by $\mathcal{V}_j^{(i)} = [a_{j-1}^{(i)}, a_j^{(i)}]$.

It is straightforward to verify that by construction the maps f and $\hat{f}_i$ satisfy

$$f(\mathcal{V}_j^{(i)}) \subset \hat{f}_i(\mathcal{V}_j^{(i)}) \subset \mathcal{V}_{j_1}^{(i-1)} \qquad (31)$$

for some j_1, which depends on j. Furthermore, it can be shown (see, e.g., [25]) that the partition elements satisfy

$$\max_j |\mathcal{V}_j^{(i)}| < d^i \qquad (32)$$

for some $d < 1$. From (31) and (32), it therefore follows that for $i > k$

$$|f^k(x) - \hat{f}_i^k(x)| < C\,d^{i-k} \qquad (33)$$

for all $x \in [0, 1]$ and for some constant C. Uniform convergence of $\hat{f}_i^k$ to f^k follows from (33). ∎

## REFERENCES

[1] A. V. Oppenheim, G. W. Wornell, S. H. Isabelle, and K. M. Cuomo, "Signal processing in the context of chaotic signals," in *Proc. Int. Conf. Acoust., Speech, Signal Processing,* 1992.

[2] A. Singer, G. Wornell, and A. Oppenheim, "Nonlinear autoregressive modeling and estimation in the presence of noise," *Digital Signal Processing,* vol. 4, pp. 207–221, Oct. 1994.

[3] H. C. Papadopoulos and G. W. Wornell, "Maximum likelihood estimation of a class of chaotic signals," *IEEE Trans. Inform. Theory,* vol. 41, pp. 312–317, Jan. 1995.

[4] M. D. Richard, "Estimation and detection with chaotic systems," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, Feb. 1994.

[5] S. Haykin and X. B. Li, "Detection of signals in chaos," *Proc. IEEE,* vol. 83, pp. 95–122, Jan. 1995.

[6] S. H. Isabelle, "A signal processing framework for the analysis and application of chaos," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, Feb. 1995.

[7] B. Chen and G. W. Wornell, "Efficient channel coding for analog sources using chaotic systems," in *Proc. IEEE GLOBECOM,* Nov. 1996.

[8] L. Risbo, "On the design of tone-free sigma-delta modulators," *IEEE Trans. Circuits Syst. II,* vol. 42, pp. 52–55, Jan. 1995.

[9] J. H. B. Deane and D. C. Hamill, "Chaotic behavior in a current-mode controlled DC-DC converter," *Electron. Lett.,* vol. 27, pp. 1172–1173, 1991.

[10] C. Chase, J. Serrano, and P. J. Ramadge, "Periodicity and chaos from switched flow systems: Contrasting examples of discretely controlled continuous systems," *IEEE Trans. Automat. Contr.,* vol. 38, pp. 71–83, 1993.

[11] L. O. Chua, Y. Yao, and Q. Yang, "Generating randomness from chaos and constructing chaos with desired randomness," *Int. J. Circuit Theory Applications,* vol. 18, pp. 215–240, 1990.

[12] K. Peterson, *Ergodic Theory.* Cambridge, U.K.: Cambridge Univ. Press, 1983.

[13] I. P. Natanson, *Theory of Functions of a Real Variable.* New York: Frederick Ungar, 1961.

[14] Y. S. Tang, A. I. Mees, and L. O. Chua, "Synchronization and chaos," *IEEE Trans. Circuits Syst.,* vol. CAS-30, pp. 620–626, Sept. 1983.

[15] A. Lasota and M. C. Mackey, *Probabilistic Properties of Deterministic Systems.* Cambridge, U.K.: Cambridge Univ. Press, 1985.

[16] P. R. Halmos, *Lectures on Ergodic Theory.* Tokyo, Japan: Math. Soc. Japan, 1956.

[17] A. Boyarsky and M. Scarowsky, "On a class of transformations which have unique absolutely continuous invariant measures," *Trans. Amer. Math. Soc.,* vol. 255, pp. 243–262, 1979.

[18] R. Kalman, "Nonlinear aspects of sampled-data control systems," in *Proc. Symp. Nonlinear Circuit Anal.,* Apr. 1956, pp. 273–313.

[19] N. Friedman and A. Boyarsky, "Matrices and eigenfunctions induced by Markov maps," *Lin. Alg. App.,* vol. 39, pp. 141–147, 1981.

[20] H. Minc, *Nonnegative Matrices.* New York: Wiley, 1988.

[21] G. Strang, *Linear Algebra and its Applications.* New York: Academic, 1980, 2nd ed.

[22] A. A. Kosyakin and E. A. Sandler, "Ergodic properties of a class of
     piecewise-smooth transformations of a segment," *Izv. VUZ Mat.*, vol.
     118, pp. 32–40, 1972.
[23] A. Boyarsky and G. Haddad, "All invariant densities of piecewise linear
     markov maps are piecewise constant," *Adv. Appl. Math.*, vol. 2, pp.
     284–289, 1981.
[24] H. Sakai and H. Tokumaru, "Autocorrelations of a certain chaos," *IEEE
     Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 588–590,
     May 1980.
[25] F. Hofbauer and G. Keller, "Ergodic properties of invariant measures for
     piecewise monotonic transformations," *Math. Z.*, vol. 180, pp. 119–140,
     1982.
[26] S. Wong, "Some metric properties of mappings of the interval," *Trans.
     Amer. Math. Soc.*, vol. 246, pp. 493–500, 1978.
[27] P. Gora and A. Boyarsky, "Compactness of invariant densities for
     families of expanding, piecewise monotonic transformations," *Can. J.
     Math.*, vol. 41, no. 5, pp. 855–869, 1989.
[28] A. Boyarsky, "A matrix method for estimating the Lyapunov exponent of
     one dimensional systems," *J. Stat. Phys.*, vol. 50, no. 1/2, pp. 213–229,
     1988.

Steven H. Isabelle received the B.E.E. and M.S. degrees from the Georgia
Institute of Technology, Atlanta, and the Ph.D. degree from the Massachusetts
Institute of Technology (MIT), Cambridge.

He has been employed AT&T Bell Laboratories, Aware Inc., and MIT and
has worked on various aspects of signal processing including speech and
image processing, channel coding, wireless communications, and nonlinear
dynamical system theory.
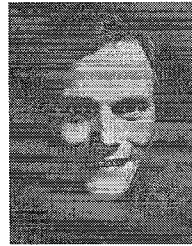
Gregory W. Wornell (M'91) was born in Montréal,
P.Q., Canada in 1962. He received the B.A.Sc.
degree (with honors) from the University of British
Columbia, Vancouver, B.C., Canada, and the S.M.
and Ph.D. degrees from the Massachusetts Institute
of Technology (MIT), Cambridge, all in electrical
engineering, in 1985, 1987 and 1991, respectively.

Since 1991, he has been on the faculty of the
Department of Electrical Engineering and Computer
Science at MIT, where he is currently Cecil and
Ida Green Career Development Associate Professor.
During the 1992–1993 academic year, he was on leave at AT&T Bell Labora-
tories, Murray Hill, NJ, and during 1990, he was a Visiting Investigator at the
Woods Hole Oceanographic Institution, Woods Hole, MA. His current research
interests include signal processing, wireless and broadband communications,
and applications of fractal geometry and nonlinear dynamical system theory in
these areas. He is author of the monograph *Signal Processing with Fractals:
A Wavelet-Based Approach* (Englewood Cliffs, NJ: Prentice-Hall).

Dr. Wornell is currently Associate Editor for the communication area for
the IEEE SIGNAL PROCESSING LETTERS and serves on the DSP Technical
Committee of the IEEE Signal Processing Society. He is also a consultant
to industry and an inventor on two patents in the area of communications;
two others are pending. Among the awards he has received for teaching and
research are the MIT Goodwin Medal for "conspicuously effective teaching"
(1991), the ITT Career Development Chair at MIT (1993), an NSF Faculty
Early Career Development Award (1995), an ONR Young Investigator Award
(1996), and the MIT Junior Bose Award for Excellence in Teaching (1996).
He is a member of Tau Beta Pi and Sigma Xi.