PROPERTIES AND DISCRIMINATION OF CHAOTIC MAPS

Michael D. Richard

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

Properties, discrimination, and practical applications of chaotic maps of the unit interval are considered. We focus on a specific class of piecewise linear, one-dimensional maps, members of which give rise to finite-state Markov chains. Properties of maps in this class are presented; these properties suggest the value of these maps for generating random variables and processes. Algorithms are introduced for discriminating among these maps based on noisy observations, and preliminary experimental results are presented which illustrate the behavior of these algorithms. We briefly speculate on the use of these maps and discrimination algorithms for communication applications.

1. INTRODUCTION

Chaotic systems have received much attention in the mathematics and physics communities in the last two decades; and they are receiving increasing attention in various engineering disciplines as well. Traditionally, researchers have focused on causes of chaos, universal properties shared by chaotic systems, as well as topological and ergodic properties of chaotic systems. To date, few practical engineering applications of chaotic systems have emerged.

This paper, in part, considers the practical value of chaotic systems. The paper shows how the distinguishing properties of a particular class of chaotic systems may render them useful for certain engineering applications, including secure communication.

A discussion of the conditions a system must satisfy to be considered chaotic is beyond the scope of the paper and is available in a number of references on the topic [3]. Of interest here is a single property of chaotic systems—the ability of these deterministic systems to generate waveforms with stochastic aspects. In addition, the paper only considers one-dimensional, discrete-time systems, or "maps", of the unit interval onto itself, and it focuses on a particular class of these maps.

The next section introduces this class of chaotic maps and discusses useful properties of maps in this class. Section 3 discusses optimal and suboptimal algorithms for discriminating among these maps based on noisy observations. Section 3 also presents preliminary experimental results obtained with these discrimination algorithms. Finally, Section 4 briefly considers two communication scenarios involving these maps and discrimination algorithms.

Throughout the paper, the term "orbit" denotes the infinite sequence of points $\{x(i)\}$ which satisfy $x(n) = f(x(n-1))$ for some chaotic map $f(\cdot)$, and the term "orbit segment" denotes a finite, consecutive set of orbit points (i.e., a piece of an orbit).

2. MARKOV MAPS

2.1. Fundamentals

In this paper, we focus on one class of chaotic maps of the unit interval—piecewise linear, chaotic, Markov maps—which for notational convenience we refer to as simply "Markov maps". Our interest in these maps arises from their many interesting properties which render them amenable to analysis and potentially useful for practical applications. We highlight several of these properties and their practical relevance in this section and the next. In addition, although Markov maps represent only a small subset of the chaotic maps of the unit interval, they closely approximate the dynamics of a much larger set of chaotic maps. In fact, for each map in this larger set, there exists a sequence of Markov maps which converges uniformly to it [2, 5].

The technical requirements a map $f(\cdot)$ must satisfy to be considered a Markov map are provided in [1]. The fundamental requirement is that one can divide the unit interval into a finite set of nonoverlapping subintervals $\{I_j\}$, for which the following two conditions hold:

1. Endpoints get mapped to endpoints. In other words, each endpoint of a subinterval is mapped by $f(\cdot)$ to an endpoint of a (possibly the same) subinterval.

2. Each subinterval is mapped "onto" a union of subintervals. That is, for each subinterval I_j, if some point in I_j gets mapped to subinterval I_k, then all points (except possibly the endpoints) in I_k are mapped to by points in I_j.

For a Markov map, a set of nonoverlapping subintervals which divide the unit interval and satisfy the above conditions is called a "Markov partition". A Markov map generally has an infinite number of Markov partitions, and two different Markov maps may have the same set of Markov partitions [7].

Figure 1 depicts two Markov maps. A Markov partition for the map on the left is given by any division of the unit interval into $2N$ equal length subintervals, where N is any positive integer. Similarly, a Markov partition for the map on the right is given by any division of the unit interval into $4N$ equal length subintervals. For example,

This work was done in the RLE Digital Signal Processing Group, MIT. This work was funded in part by the Defense Advanced Research Projects Agency monitored by ONR under Contract No. N00014-89-J-1489 and in part by the U.S. Air Force Office of Scientific Research under Grant Nos. AFOSR-91-0034-A and F49620-92-J-0255.

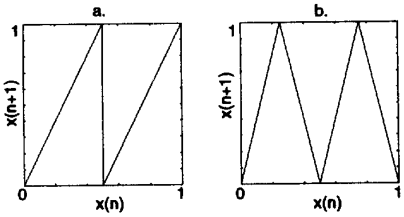

Figure 1: Two Markov Maps

the partition given by the four equal length subintervals $\{[0,.25),[.25,.5),[.5,.75),[.75,1]\}$ is a Markov partition for each map.

A distinguishing property of Markov maps is that they give rise to Markov chains [1, 4]. In particular, each element I_j of a Markov partition for a Markov map $f(\cdot)$ corresponds to a state S_j in a Markov chain corresponding to that partition. The transition probability from state S_j to state S_k equals the fraction of points in the partition element (subinterval) I_j that are mapped to the partition element (subinterval) I_k. For the Markov maps shown in Figure 1, with the Markov partition given by $\{[0,.25),[.25,.5),[.5,.75),[.75,1]\}$, the transition matrices are given by

$$\overset{a.}{\begin{bmatrix} .5 & .5 & 0 & 0 \\ 0 & 0 & .5 & .5 \\ .5 & .5 & 0 & 0 \\ 0 & 0 & .5 & .5 \end{bmatrix}} \qquad \overset{b.}{\begin{bmatrix} .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \end{bmatrix}}.$$

The dynamics of the Markov chain arise as follows. If $x(n)$, the orbit point at time n (i.e., $x(n) = f(x(n-1)) = f^n(x(0))$), is in partition element I_j, then the Markov chain is said to be in state S_j at time n. For almost all initial conditions $x(0)$, the state sequence that arises as a result of this mapping between orbit points and states is a first-order Markov process with transition probabilities defined as above.

2.2. Properties

Markov maps have many interesting, potentially useful properties. For example, as noted in [4], given the transition matrix for any finite-state Markov chain, one can construct a Markov map which gives rise to that Markov chain. That is, one can construct a Markov map which has a Markov partition for which the corresponding Markov chain has the same transition matrix as the desired transition matrix. Thus, Markov maps are useful generators of arbitrary, finite-state Markov chains.

A well-known result in ergodic theory is that every ergodic, finite-state Markov chain has a unique, invariant state probability vector. Similarly, under fairly mild conditions, a Markov map has an invariant probability density function, and if it exists, this density function is piecewise constant. As one might expect, a close relation exists between the invariant probability density function of a Markov map and the invariant probability vectors of Markov chains it gives rise to. A potentially useful result is that given any valid, piecewise constant, probability density function (with a finite number of pieces), one can easily construct a Markov map that has this function as its invariant density function [7]. For each of the two Markov maps shown in Figure 1, the invariant density is simply the constant value 1 over the unit interval. The two Markov maps shown on the left in Figure 2 have nontrivial invariant density functions. The

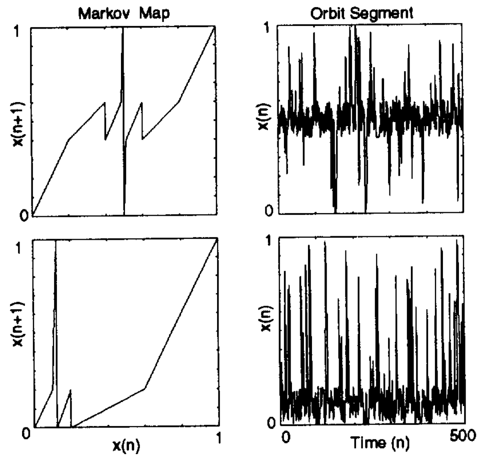

Figure 2: Markov Maps with Nontrivial Invariant Densities

invariant density for the top map equals 4 over the subinterval $[.4,.6)$ and .25 elsewhere, whereas the invariant density for the bottom map equals 4 over the subinterval $[.0,.2)$ and .25 elsewhere. The 500-point orbit segment shown to the right of each map graphically indicates the concentration of orbit points in the appropriate subinterval with invariant density equal to 4. This relation between piecewise constant, invariant probability densities and Markov maps suggests that these simple maps may be useful generators of random variables with density functions that approximate arbitrary probability density functions.

3. DISCRIMINATING AMONG MARKOV MAPS

3.1. Problem Scenario

The close relation between Markov maps and Markov chains allows one to derive simple, computationally efficient algorithms for discriminating among these maps based on noisy observations. This section discusses discrimination algorithms for two special cases of the following, general discrimination scenario.

One is given M known Markov maps, $\{f_i(\cdot)\}_{i=1}^M$, and a sequence of N observations $Y = \{y(i)\}_{i=1}^N$, where each observation satisfies

$$y(n) = h_k(x(n)) + v(n), \quad 1 \le n < N. \tag{1}$$

In this equation, $X = \{x(i)\}_{i=1}^N$ is an orbit segment from one of the Markov maps $f_k(\cdot)$ (i.e., $x(n) = f_k(x(n-1))$) and $\{v(i)\}_{i=1}^N$ is a white-noise sequence with known statistics, which is assumed to be independent of the initial condition $x(0)$ and the chosen map $f_k(\cdot)$. Also, $h_k(\cdot)$ is a memoryless transformation possibly dependent on the map $f_k(\cdot)$. In Section 3.2, we consider the case in which $h_k(\cdot)$ is a quantizer, and in Section 3.3, the case in which $h_k(\cdot)$ is the identity operator.

The discrimination task is to use the observations Y to determine which of the M maps generated the unobserved orbit segment X which gave rise to these observations. A fundamental result from estimation theory is that for equally likely maps, the optimal discrimination rule (in a minimum probability of error or maximum likelihood sense) is to choose that map among the M, with the largest likelihood $p(Y|f_k(\cdot))$, where $p(Y|f_k(\cdot))$ is the probability density of the observations Y, given that the map $f_k(\cdot)$ generated the orbit segment X giving rise to these observations.

3.2. Quantized orbit points

A computationally efficient implementation of the optimal discrimination rule exists when each of the M transformations $h_k(\cdot)$ is a quantizer which associates a single, unique value with each element of a Markov partition for $f_k(\cdot)$. That is, if I_j^k denotes the j^{th} partition element for map $f_k(\cdot)$ and H_j^k denotes the value associated with this partition element by $h_k(\cdot)$, then $h_k(x) = H_j^k$ for all $x \in I_j^k$. (Typical values of H_j^k include the midpoint or either endpoint of I_j^k). With the M quantizers chosen this way, the discrimination problem reduces to that of discriminating among M hidden Markov models (HMMs). As a result, one can use the "forward" portion of the computationally efficient "forward-backward" algorithm [6] to calculate the exact values of the M likelihoods $p(Y|f_k(\cdot))$ used in the optimal discrimination rule.

Specifically, for each map $f_k(\cdot)$, the partition elements $\{I_j^k\}_{j=1}^{T_k}$ correspond to the "unobserved" states in the hidden Markov model associated with that map, where T_k denotes the number of partition elements for the Markov partition associated with the map. For the discrimination scenario introduced in Section 3.1, the output at time n, $o_j^k(n)$, associated with I_j^k is given by

$$o_j^k(n) = H_j^k + v(n). \tag{2}$$

In other words, if $f_k(\cdot)$ generated the orbit segment $X = \{x(i)\}_{i=1}^N$ and $x(n) \in I_j^k$, then $y(n)$, the observation at time n, is given by

$$y(n) = h_k(x(n)) + v(n) = o_j^k(n) = H_j^k + v(n). \tag{3}$$

With this observation equation, $p(y(n)|x(n) \in I_j^k, f_k(\cdot))$, which is the probability density of the observation $y(n)$ conditioned on the map $f_k(\cdot)$ having generated the orbit segment and $x(n)$ being in partition element I_j^k, is given by

$$p(y(n)|x(n) \in I_j^k, f_k(\cdot)) = p_v(v(n) = y(n) - H_j^k) \tag{4}$$

where $p_v(\cdot)$ is the density function associated with the white-noise sequence.

Next, define the "forward variable" $\alpha_j^k(n)$ as

$$\alpha_j^k(n) = p(y(1), y(2), \cdots, y(n), x(n) \in I_j^k | f_k(\cdot)). \tag{5}$$

Thus, $\alpha_j^k(n)$ is the joint probability density of the observations through time n and $x(n) \in I_j^k$, conditioned on map $f_k(\cdot)$ having generated the orbit segment $X = \{x_i\}_{i=1}^N$. Note that $\alpha_j^k(n)$ can also be expressed

$$\begin{aligned} \alpha_j^k(n) &= p(y(1), y(2), \cdots, y(n)|x(n) \in I_j^k, f_k(\cdot)) \\ &\quad \times p(x(n) \in I_j^k | f_k(\cdot)). \end{aligned} \tag{6}$$

Finally, let $\{s_{l,j}^k\}_{j,l=1}^{T_k}$ denote the state transition probabilities associated with $f_k(\cdot)$, where $s_{l,j}^k$ denotes the probability that $x(n) \in I_j^k$ given that $x(n-1) \in I_l^k$. With these definitions and notational conventions, an efficient implementation of the optimal discrimination rule is the following:

1. For each map $f_k(\cdot)$, compute $\alpha_j^k(n)$ as a function of n for $j = 1, \cdots, T_k$ by using the following recursion:

$$\begin{aligned} \alpha_j^k(1) &= p(y(1)|x(1) \in I_j^k, f_k(\cdot)) \\ &\quad \times p(x(1) \in I_j^k | f_k(\cdot)) \end{aligned} \tag{7}$$

$$\begin{aligned} \alpha_j^k(n+1) &= \left[\sum_{l=1}^{T_k} \alpha_l^k(n) s_{l,j}^k\right] \\ &\quad \times p(y(n+1)|x(n+1) \in I_j^k, f_k(\cdot)) \end{aligned} \tag{8}$$

2. For each map, compute the likelihood $p(Y|f_k(\cdot))$ by exploiting the relation given by

$$p(Y|f_k(\cdot)) = \sum_{j=1}^{T_k} \alpha_j^k(N). \tag{9}$$

3. Choose the map $f_k(\cdot)$ for which $p(Y|f_k(\cdot))$ is largest.

Note that the above discrimination rule requires specification of initial state probabilities $p(x(1) \in I_j^k|f_k(\cdot))$ for each map.

3.3. Unquantized orbit points

When each of the M transformations $h_k(\cdot)$ is the identity operator, the discrimination problem is that of discriminating among M Markov maps based on noisy observations of unquantized orbit points. In general, for the case of unquantized orbit points, optimal discrimination among the M Markov maps is not computationally feasible because the initial condition $x(0)$ is unknown and the maps are chaotic. However, if the Markov maps each allow arbitrarily fine Markov partitions, we can perform robust, computationally efficient, suboptimal discrimination by modeling the dynamics of each map as a hidden Markov model and applying the discrimination rule outlined in Section 3.2.

Specifically, we first find a "sufficiently fine" Markov partition for each map. (A discussion of how to quantify the nebulous term "sufficiently fine" is available in [7] as is a discussion of the conditions under which a Markov map allows arbitrarily fine Markov partitions.) Having found these Markov partitions, we apply the same discrimination rule as outlined in Section 3.2, with one important modification. In particular, we replace the expression for the output at time n, $o_j^k(n)$, given by (2) with the following

$$o_j^k(n) = u_j^k(n) + v(n), \tag{10}$$

where $u_j^k(n)$ is a uniform random variable independent of $v(n)$ with region of support over I_j^k. Thus, the observation $y(n)$ given by (3) is now the sum of two independent random variables with conditional density function $p(y(n)|x(n) \in I_j^k, f_k(\cdot))$ given by the convolution of the density functions for $u_j^k(n)$ and $v(n)$.

This HMM-based approach to discrimination is useful even for maps which are not Markov maps, but instead belong to the class of chaotic maps for which there are uniformly converging sequences of Markov maps. For each of these maps, one would first find an approximating Markov map and then apply the above approach to these approximating maps. In addition, one can use the HMM-based modeling approach to derive computationally efficient algorithms for performing "approximate" Maximum Likelihood and Bayesian state estimation with Markov maps and with chaotic maps that are well-approximated by Markov maps [7].

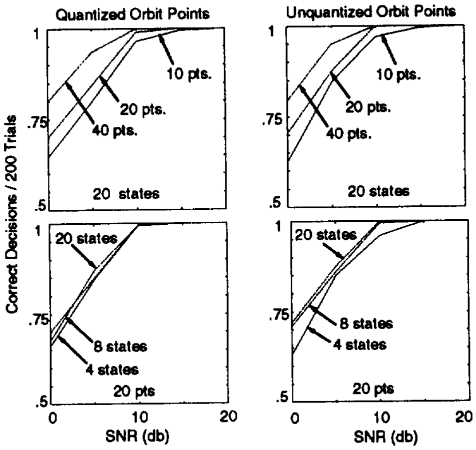

Figure 3: Discrimination Results with Markov Maps Shown in Figure 1

3.4. Computer Experiments

Figure 3 depicts preliminary results for discriminating between the Markov maps shown in Figure 1. Gaussian white noise was used for the white-noise sequence $\{v(i)\}_{i=1}^{N}$ in (1) when generating the observations Y. The results with quantized orbit points were obtained by using the left endpoint of each partition element as the value associated with that partition element by the quantizer. In addition, the ratio used to determine the signal-to-noise ratio (SNR) was that of the variance of the orbit generated by each map divided by the variance of the white-noise. Since each map in Figure 1 has a uniform invariant density, the variance of the orbit generated by each map (with unquantized orbit points) is the same as that of a uniform random variable with region of support over the unit interval. (In general, the variance of an orbit segment with quantized points differs slightly from the variance of an orbit segment with unquantized points.)

The results shown in the left two graphs were obtained by applying the optimal discrimination rule to quantized orbit points. The results shown in the right two graphs were obtained by applying the suboptimal discrimination rule to unquantized orbit points. Each plotted result represents the fraction of correct discrimination decisions for 200 independent trials, with each map being the correct map for 100 trials. In light of the small number of trials used, the plotted results should not be interpreted as probabilities, but as qualitative indicators of the behavior of the discrimination algorithms on the two maps used in the experiments.

The curves in the top graphs are parameterized by the number of noise-corrupted observation points used for discrimination, and all results in these graphs were obtained with a Markov partition consisting of 20 equally sized subintervals. In contrast, the curves in the bottom figures are parameterized by the number of equally sized subintervals comprising the Markov partition, and all results were obtained with 20 noise-corrupted observations points. Not surprisingly, discrimination performance improves as the number of observations increases for a given signal-to-noise ratio (SNR). What is surprising is the comparable performance achieved with both quantized and unquantized orbit points. Also, the results suggest that discrimination performance is relatively insensitive to the size of the Markov partition. As shown in [7], this is not a universal result and with some Markov maps there is a strong correlation between discrimination performance and the size of the Markov partition.

4. POTENTIAL APPLICATIONS

In light of their various properties, Markov maps and chaotic maps well-approximated by them may be useful for secure M-ary communication applications. For example, one could associate a unique Markov map to each of the M signals. To transmit a signal, one would transmit a (possibly quantized) fixed-length orbit segment of the corresponding map. A different orbit segment would be used each time that signal was transmitted. Assuming an additive, white-noise channel, one would apply the discrimination algorithms discussed in Section 3 at the receiver to decide which signal was sent. Note that for a stream of transmitted signals (i.e., orbit segments), at the receiver one would not need to determine the exact starting and ending locations of each orbit segment, but would only need to isolate a "sufficiently large" subsegment of each segment. This might offer the approach an advantage over traditional direct-sequence spread spectrum techniques, since they require knowledge, at the receiver, of the precise starting and ending locations of each transmitted signal.

Alternatively, instead of transmitting fixed-length orbit segments, one could transmit variable-length orbit segments. In this case, at the receiver one would need to determine the signal corresponding to each transmitted orbit segment as before. However, one would also now need to determine the approximate transition locations among transmitted orbit segments. If the orbit points were properly quantized, one could model the problem of determining these transitions as one of detecting instants of changes between random processes, with each process being an HMM. Because of its recursive structure, the "forward algorithm" could be incorporated in a computationally efficient thresholding scheme for detecting these instants of changes.

REFERENCES

[1] A. Boyarsky and M. Scarowsky, "On a Class of Transformations Which Have Unique Absolutely Continuous Invariant Measures," *Transactions of the American Mathematical Society*, Vol. 255, November 1979, pp. 243–262.

[2] A. Boyarsky, "A Matrix Method for Estimating the Liapunov Exponent of One-Dimensional Systems," *Journal of Statistical Physics*, Vol. 50, 1988, pp. 213–229.

[3] R. Devaney, *An Introduction to Chaotic Dynamical Systems*. Redwood City: Addison-Wesley, 1989.

[4] R. Kalman, "Nonlinear aspects of sampled-data control systems," in *Proceedings of the Symposium on Nonlinear Circuit Analysis*, Polytechnic Institute of Brooklyn, 25-27 April 1956, pp. 273–313.

[5] A. Kosyakin and E. Sandler, *IZV Matematika*, Vol. 3, No. 118, 1972, pp. 32–40.

[6] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications," *Proceedings of the IEEE*, Vol. 77, No. 2, February 1989, pp. 257–286.

[7] M. Richard, Ph.D. Thesis, in preparation.