Convergence of eigenvalues in state-discretization of linear stochastic systems

Jose A. de Dona
University of Newcastle

Graham C. Goodwin
University of Newcastle

Richard H. Middleton
University of Newcastle

Iain Raeburn
University of Wollongong, raeburn@uow.edu.au

Convergence of eigenvalues in state-discretization of linear stochastic systems

Abstract

The transition operator that describes the time evolution of the state probability distribution for continuous-state linear systems is given by an integral operator. A state-discretization approach is proposed, which consists of a finite rank approximation of this integral operator. As a result of the state-discretization procedure, a Markov chain is obtained, in which case the transition operator is represented by a transition matrix. Spectral properties of the integral operator for the continuous-state case are presented. The relationships between the integral operator and the finite rank approximation are explored. In particular, the limiting properties of the eigenvalues of the transition matrices of the resulting Markov chains are studied in connection to the eigenvalues of the original continuous-state integral operator.

Keywords

linear, stochastic, discretization, state, convergence, systems, eigenvalues

Disciplines

Engineering | Science and Technology Studies

Publication Details

CONVERGENCE OF EIGENVALUES IN STATE-DISCRETIZATION OF LINEAR STOCHASTIC SYSTEMS*

JOSÉ A. DE DONÁ[†], GRAHAM C. GOODWIN[†], RICHARD H. MIDDLETON[†], AND IAIN RAEBURN[‡]

Abstract. The transition operator that describes the time evolution of the state probability distribution for continuous-state linear systems is given by an integral operator. A state-discretization approach is proposed, which consists of a finite rank approximation of this integral operator. As a result of the state-discretization procedure, a Markov chain is obtained, in which case the transition operator is represented by a transition matrix. Spectral properties of the integral operator for the continuous-state case are presented. The relationships between the integral operator and the finite rank approximation are explored. In particular, the limiting properties of the eigenvalues of the transition matrices of the resulting Markov chains are studied in connection to the eigenvalues of the original continuous-state integral operator.

Key words. state-discretization, Markov chains, integral operators, finite rank approximation

AMS subject classifications. 60J10, 47N30

PII. S0895479897329874

1. Introduction. The understanding and use of links between continuous and discrete features of systems has proven to be a fruitful area. In some cases, knowledge of properties from discrete processes gives insight into the continuous counterpart. On the other hand, the mathematical methods associated with the continuum (e.g., differential equations, integrals) very often turn out to be easier for analytical treatment than those associated with discrete coordinate axes (e.g., difference equations, summations) [2].

The algorithms available for solving discrete-state problems, for example, parameter estimation in Markov chains, have a very high computational complexity when the number of states is large. Typically the number of computations required is proportional to the square of the number of states, for example, the Viterbi algorithm or the probability grid filter (see, e.g., [13]). On the other hand, the tools available for solving linear continuous-state problems (e.g., the celebrated Kalman filter) are well known for their simplicity.

With these advantages of continuous-state algorithms over discrete-state ones in mind, we are interested in looking at continuous-state approximations to discrete-state problems. This line of research is aimed at identifying the appropriate geometric structures in discrete-state models which are amenable to linear continuous-state approximations. As part of this program of characterization of the class of discrete-state systems suitable for continuous-state approximation, we are studying the inverse process,

namely, the properties of discrete-state systems that arise in the state-discretization of continuous-state linear systems. In particular, this paper examines the asymptotic eigenstructure of the discrete-state systems (Markov chains) generated by the state-discretization of linear continuous-state systems. Related work can be found in [4, 6, 10] in a general setting. However, since we are restricting our study to linear continuous-state systems, we are able to provide more specific information about the structure of the discrete-state approximation.

This paper is organized as follows. In section 2 we present some properties of state space formulations of linear systems that are relevant for their state-discretization. An operator which provides the time evolution of continuous-state probability density functions, together with its eigenvalues and eigenfunctions, is discussed in that section. In section 3 a state-discretization approach for linear systems is presented, and properties of the resulting discrete-state Markov chains are discussed. In particular, it is proved that the eigenvalues of the transition matrices obtained converge to a subset of the eigenvalues of the original continuous-state operator. Section 4 is dedicated to an example in order to illustrate the material presented.

2. Discrete-time linear state space models. Let a linear discrete-time system be described by the state space equation

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{v}_k, \tag{1}$$

where $\mathbf{x}_k, \mathbf{x}_{k+1} \in \mathbb{R}^n$ are the states at discrete-times k and $k+1$, respectively, $\mathbf{v}_k \in \mathbb{R}^m$ is an independent and identically distributed (iid) sequence of random vectors with finite variance, and the initial state $\mathbf{x}_0$ is assumed to have an arbitrary distribution with finite variance and independent of $\mathbf{v}_k$ for all k. $\mathbf{A}$ and $\mathbf{B}$ are real matrices of appropriate dimension, and we assume $\mathbf{A}$ such that all its eigenvalues belong to the interior of the unit disk. For simplicity of presentation we will further assume that the matrix $\mathbf{A}$ is invertible (see, e.g., (2)). This condition is naturally satisfied in the context of time-discretization of a continuous time system (see, e.g., [1]).

We will denote the joint probability density function (PDF) of a vector of n random variables $\mathbf{y}$, evaluated at the n-tuple $(x_1, x_2, \ldots, x_n)$, by $f_{\mathbf{y}}(\mathbf{x}) \triangleq f_{\mathbf{y}}(x_1, x_2, \ldots, x_n)$, where $\mathbf{x}$ denotes the vector $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$.

We assume here that all the PDFs of interest have compact support and belong to the Hilbert space $\mathcal{L}^2([a_1, b_1] \times \cdots \times [a_n, b_n])$ (e.g., the uniform distribution). This assumption is necessary in the proofs below, and from a computational point of view, in most interesting applications densities do not significantly differ from zero outside a suitable compact set.

The PDF of $\mathbf{x}_{k+1}$ can then be expressed in terms of those of $\mathbf{x}_k$ and $\mathbf{v}_k$ (for example, by using the auxiliary variables method; see, e.g., [8]) as follows:

$$
\begin{aligned}
f_{\mathbf{x}_{k+1}}(\mathbf{x}) &= \frac{1}{|\det(\mathbf{A})|} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\mathbf{x}_k}\left(\mathbf{A}^{-1}(\mathbf{x} - \mathbf{B}\boldsymbol{\xi})\right) f_{\mathbf{v}}(\boldsymbol{\xi}) d\xi_1 \ldots d\xi_m \\
&\triangleq (Tf_{\mathbf{x}_k})(\mathbf{x}),
\end{aligned}
\tag{2}
$$

where $\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}^m$ and (2) defines a convex operator on the convex space of PDFs. Since we are interested in spectral properties, we work with the extension of this operator to the linear space $\mathcal{L}^2([a_1, b_1] \times \cdots \times [a_n, b_n])$. Note that this extension is the linear operator T on $\mathcal{L}^2([a_1, b_1] \times \cdots \times [a_n, b_n])$ defined by (2). In (2) we omitted k in $f_{\mathbf{v}_k}$ since $\mathbf{v}_k$ is iid.

In what follows we use the *n-dimensional characteristic function* (which exists for all distributions; see, e.g., [9]) and denote

$$\Phi_{\mathbf{x}_k}(i\mathbf{w}) = \mathcal{F}\{f_{\mathbf{x}_k}(\mathbf{x})\} = E\exp\{-i\mathbf{w}^T\mathbf{x}_k\}, \tag{3}$$

$$\Phi_{\mathbf{v}}(i\mathbf{w}) = \mathcal{F}\{f_{\mathbf{v}}(\mathbf{x})\} = E\exp\{-i\mathbf{w}^T\mathbf{v}\}, \tag{4}$$

where $\mathcal{F}\{\cdot\}$ stands for the Fourier transform and $i \triangleq \sqrt{-1}$. In expression (3) $\mathbf{w} \in \mathbb{R}^n$, and in (4) $\mathbf{w} \in \mathbb{R}^m$.

In the Fourier transform domain (2) becomes

$$\Phi_{\mathbf{x}_{k+1}}(i\mathbf{w}) = \Phi_{\mathbf{x}_k}(\mathbf{A}^T i\mathbf{w})\Phi_{\mathbf{v}}(\mathbf{B}^T i\mathbf{w}) \triangleq (\tilde{T}\Phi_{\mathbf{x}_k})(i\mathbf{w}), \tag{5}$$

where $\mathbf{w} = (w_1, \dots, w_n)^T \in \mathbb{R}^n$ and $\tilde{T}$ is the linear operator on Fourier transforms (which corresponds to the linear operator T).

Since $\mathbf{v}_k \in \mathbb{R}^m$ is iid and the eigenvalues of matrix $\mathbf{A}$ are less than unity in modulus, there exists a *stationary PDF*, denoted $f_{\mathbf{x}_\infty}$, such that satisfies $f_{\mathbf{x}_\infty} = Tf_{\mathbf{x}_\infty}$ in (2) (see, for example, Theorems 2.3 and 2.7 in [7] for a proof of the more general case where the coefficients in the autoregressive model are random).

We let $\Phi_{\mathbf{x}_\infty}$ denote the Fourier transform of the stationary distribution, satisfying

$$\Phi_{\mathbf{x}_\infty}(i\mathbf{w}) = (\tilde{T}\Phi_{\mathbf{x}_\infty})(i\mathbf{w}) = \Phi_{\mathbf{x}_\infty}(\mathbf{A}^T i\mathbf{w})\Phi_{\mathbf{v}}(\mathbf{B}^T i\mathbf{w}). \tag{6}$$

For simplicity, the following theorem is formulated in the Fourier domain, although an equivalent result can be obtained in the state domain (see Remark 2.1).

THEOREM 2.1. *The complex numbers $\lambda_{\tilde{T}}$ and the functions $V_{\tilde{T}}$ given by*

$$\lambda_{\tilde{T}} = (\lambda_1)^{l_1}(\lambda_2)^{l_2} \dots (\lambda_r)^{l_r}, \tag{7}$$

$$V_{\tilde{T}}(i\mathbf{w}) = (\mathbf{v}_1^T i\mathbf{w})^{l_1}(\mathbf{v}_2^T i\mathbf{w})^{l_2} \dots (\mathbf{v}_r^T i\mathbf{w})^{l_r}\Phi_{\mathbf{x}_\infty}(i\mathbf{w}) \tag{8}$$

are eigenvalues and eigenfunctions of the operator $\tilde{T}$, where $\lambda_1, \dots, \lambda_r$ are the eigenvalues corresponding to each block of the Jordan canonical form of matrix $\mathbf{A}$, $r \leq n$ is the number of blocks, $\mathbf{v}_1, \dots, \mathbf{v}_r$ are the corresponding eigenvectors, and $l_1, \dots, lr$ are arbitrary nonnegative integers.

Proof. Substituting expression (8) into (5), we get

$$(\tilde{T}V_{\tilde{T}})(i\mathbf{w}) \triangleq V_{\tilde{T}}(\mathbf{A}^T i\mathbf{w})\Phi_{\mathbf{v}}(\mathbf{B}^T i\mathbf{w})$$
$$= (\mathbf{v}_1^T \mathbf{A}^T i\mathbf{w})^{l_1}(\mathbf{v}_2^T \mathbf{A}^T i\mathbf{w})^{l_2} \dots (\mathbf{v}_r^T \mathbf{A}^T i\mathbf{w})^{l_r}\Phi_{\mathbf{x}_\infty}(\mathbf{A}^T i\mathbf{w})\Phi_{\mathbf{v}}(\mathbf{B}^T i\mathbf{w}).$$

Now, by using (6) and the fact that λ_i and $\mathbf{v}_i$ are eigenvalues and eigenvectors of $\mathbf{A}$, we have

$$(\tilde{T}V_{\tilde{T}})(i\mathbf{w}) = (\lambda_1)^{l_1}(\lambda_2)^{l_2} \dots (\lambda_r)^{l_r}(\mathbf{v}_1^T i\mathbf{w})^{l_1}(\mathbf{v}_2^T i\mathbf{w})^{l_2} \dots (\mathbf{v}_r^T i\mathbf{w})^{l_r}\Phi_{\mathbf{x}_\infty}(i\mathbf{w}),$$

which, from (7) and (8), can be written as

$$(\tilde{T}V_{\tilde{T}})(i\mathbf{w}) = \lambda_{\tilde{T}}V_{\tilde{T}}(i\mathbf{w}). \qquad \Box$$

Remark 2.1. The eigenvalues of $\tilde{T}$ given by (7) are also eigenvalues of T (denoted λ_T). The eigenfunctions of T can be evaluated by noting that $(iw_j)^l\Phi_{\mathbf{x}_\infty}(i\mathbf{w})$ is the Fourier transform of $\partial^l f_{\mathbf{x}_\infty}(\mathbf{x})/\partial x_j^l$ (see, e.g., [14]). Hence, from (8) it can be seen that the eigenfunctions of T corresponding to the eigenvalues λ_T are linear combinations of high order partial derivatives of $f_{\mathbf{x}_\infty}(\mathbf{x})$ with respect to the components of $\mathbf{x} = (x_1, \dots, x_n)^T$.

3. State-discretization of linear state space models. The linear operator T defined by (2) provides the time evolution of the continuous-state probability density functions. The state-discretization of the system (1) is performed by approximating this operator and the PDFs by discrete-state counterparts. We will restrict our study to a bounded region of the state space: an n-dimensional hypercube $\mathcal{X} = [a_1, b_1] \times \cdots \times [a_n, b_n]$ of volume $V_\mathcal{X}$. As explained before, we assume that all the PDFs of interest have compact support inside this region and belong to $\mathcal{L}^2([a_1, b_1] \times \cdots \times [a_n, b_n])$.

To discretize the region $\mathcal{X}$ of the state space we use the idea of *refinements* [6, 10], i.e., we divide $\mathcal{X}$ into N^n subsets of size $V_\mathcal{X}/N^n$ denoted $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{N^n}$, and define a Markov chain on the resulting discrete state space.

The probability vectors of the Markov chain are given by

$$\text{(9a)} \qquad \mathbf{p}_k \triangleq (P\{\mathbf{x}_k \in \mathbf{e}_1\}, \ldots, P\{\mathbf{x}_k \in \mathbf{e}_{N^n}\})^T,$$

where

$$\text{(9b)} \qquad P\{\mathbf{x}_k \in \mathbf{e}_j\} = \int \cdots \int_{\mathbf{e}_j} f_{\mathbf{x}_k}(\mathbf{x}) dx_1 \ldots dx_n,$$

and the conditional PDF of $\mathbf{x}_{k+1}$ is computed according to (2) as

$$\text{(10)} \qquad f_{\mathbf{x}_{k+1}}(\mathbf{x}|\mathbf{x}_k \in \mathbf{e}_j) = \frac{1}{|\det(\mathbf{A})|} \int_{-\infty}^{+\infty} \cdots$$
$$\cdots \int_{-\infty}^{+\infty} f_{\mathbf{x}_k}\left(\mathbf{A}^{-1}(\mathbf{x} - \mathbf{B}\boldsymbol{\xi})|\mathbf{x}_k \in \mathbf{e}_j\right) f_{\mathbf{v}}(\boldsymbol{\xi}) d\xi_1 \ldots d\xi_m.$$

In (10), $f_{\mathbf{x}_k}(\,\cdot\,|\mathbf{x}_k \in \mathbf{e}_j)$ is the conditional PDF of the random variable $\mathbf{x}_k$ evaluated at $(\cdot)$, given the condition that $\mathbf{x}_k$ belongs to the set $\mathbf{e}_j$ and is defined as (see, e.g., [8])

$$\text{(11)} \qquad f_{\mathbf{x}_k}(\mathbf{y}|\mathbf{x}_k \in \mathbf{e}_j) = \begin{cases} \dfrac{f_{\mathbf{x}_k}(\mathbf{y})}{\int \cdots \int_{\mathbf{e}_j} f_{\mathbf{x}_k}(\mathbf{x}) dx_1 \ldots dx_n} & \text{if } \mathbf{y} \in \mathbf{e}_j, \\[2ex] 0 & \text{otherwise.} \end{cases}$$

The transition matrix between time k and $k+1$, $\mathbf{Q}_k = \{q_{ij}^{(k)}\}$, $1 \le i, j \le N^n$, is then defined as

$$\text{(12)} \qquad q_{ij}^{(k)} \triangleq P\{\mathbf{x}_{k+1} \in \mathbf{e}_i | \mathbf{x}_k \in \mathbf{e}_j\} = \int \cdots \int_{\mathbf{e}_i} f_{\mathbf{x}_{k+1}}(\mathbf{x}|\mathbf{x}_k \in \mathbf{e}_j) dx_1 \ldots dx_n.$$

Notice that we have made explicit the dependence of $\mathbf{Q}_k$ on k. If this is not the case, we say that the Markov chain has stationary or homogeneous transition probabilities, and we denote by $\mathbf{Q}$ the transition matrix. Notice also that the matrix $\mathbf{Q}_k$ is defined by (10) and (12), and these depend on the PDF $f_{\mathbf{x}_k}$ (see (11)). In order to simplify the computation of the matrix $\mathbf{Q}_k$, we approximate the PDFs $f_{\mathbf{x}_k}$ used in (11) by functions that assume piecewise-constant values in each subset $\mathbf{e}_i$ (see the discussion preceding (24)). This approach has the advantage that the transition matrices are stationary (see the independence from $f_{\mathbf{x}_k}$ in the elements q_{ij} of $\mathbf{Q}$ in (24)).

We denote by λ_T the eigenvalues of the operator T and by $\lambda_{\mathbf{Q}}(N)$ the eigenvalues of the transition matrix $\mathbf{Q}$ obtained with a partition of N^n cells. Notice that the spectrum of $\mathbf{Q}$, $\sigma(\mathbf{Q})$, has N^n points. We are interested in the accumulation (or limit) points of these sets as $N \to \infty$. The following theorem provides a connection between these accumulation points and the eigenvalues of the operator T.

THEOREM 3.1. *The set of nonzero limit points, as $N \to \infty$, of eigenvalues $\lambda_{\mathbf{Q}}(N)$ is a subset of the set of eigenvalues λ_T of the operator T.*

Proof. To maintain the notation as simple as possible, we will prove the theorem for the case of scalar systems ($n = 1$). The proof for the multidimensional case is similar, but with far more complicated notation.

We consider a scalar linear system described by

$$(13) \qquad x_{k+1} = ax_k + bv_k,$$

where $0 \neq |a| < 1$ and $b \neq 0$. As explained before, it is assumed that all distributions have compact support. In particular we suppose that supp $f_v \subseteq [-B, B]$ is fixed.

In the scalar case, the operator T defined by (2) can be written as

$$(Tf_{x_k})(x) = \int_{-\infty}^{+\infty} k(x, y) f_{x_k}(y) dy,$$

where the kernel is given by $k(x, y) = f_v(\frac{x-ay}{b})/|b|$ and satisfies $\int_{-\infty}^{+\infty} k(x, y) dx = 1$. Here we choose $A \geq |b|B/(1 - |a|)$ to ensure that the support of Tf_{x_k} lies in $[-A, A]$ whenever supp $f_{x_k} \subseteq [-A, A]$ (and, in a practical situation, we also choose A large enough to ensure that the support of the initial state distribution satisfies supp $f_{x_0} \subseteq [-A, A]$). Then T becomes the integral operator on $\mathcal{L}^2([-A, A])$:

$$(14) \qquad (Tf)(x) = \int_{-A}^{A} k(x, y) f(y) dy$$

with $k \in \mathcal{L}^2([-A, A] \times [-A, A])$.

To perform the state-discretization of this system we divide the interval $\mathcal{X} = [-A, A]$ into $N^1 = N$ subsets of length $\frac{2A}{N}$, denoted $e_1, \dots, e_N$, and we define the operator P_N on $\mathcal{L}^2([-A, A])$:

$$(15) \qquad P_N \triangleq \sum_{j=1}^{N} I_{N,j},$$

where the operators $I_{N,j}$ are defined as follows:

$$(16) \qquad (I_{N,j}f)(x) = \begin{cases} \frac{N}{2A} \int_{e_j} f(z) dz & \text{if } x \in e_j, \\ 0 & \text{otherwise} \end{cases}$$

for $f \in \mathcal{L}^2([-A, A])$.

Notice that the operator P_N is a projection onto the N-dimensional subspace of $\mathcal{L}^2([-A, A])$ consisting of functions with N piecewise constant values, and that $P_N f = E[f(U)|\mathcal{G}_N]$, where $\mathcal{G}_N$ is the sigma-algebra generated by the e_i, U is the identity mapping, and the probability space $(\Omega, \mathcal{F}, P)$ is such that $\Omega = \mathbb{R}$ and P is the uniform probability measure on $[-A, A]$.

Direct computation shows that the image of the operator $P_N T P_N$ is constant over each cell e_i, taking the value

$$(17) \qquad (P_N T P_N f)(x) = \frac{N}{2A} \sum_{j=1}^{N} \left[\left(\int_{e_i} \int_{e_j} k(x,y) dy dx \right) \left(\frac{N}{2A} \int_{e_j} f(z) dz \right) \right]$$

for $x \in e_i$.

We need to prove now that the operator $P_N T P_N$ converges, as $N \to \infty$, to T in the natural norm of operators on $\mathcal{L}^2([-A, A])$. For this purpose we use $k_{i,j}$ to denote the scaled integral of $k(x,y)$ over the cell $e_i \times e_j$:

$$(18) \qquad k_{i,j} = \left(\frac{N}{2A} \right)^2 \int_{e_i} \int_{e_j} k(x,y) dy dx,$$

and we define a *piecewise-constant* function on $[-A, A] \times [-A, A]$ by

$$(19) \qquad n_N^k(x,y) = \sum_{i=1}^{N} \sum_{j=1}^{N} k_{i,j} \mathcal{X}_{e_i}(x) \mathcal{X}_{e_j}(y),$$

where $\mathcal{X}_{e_i}(\cdot)$ is the indicator function $\mathcal{X}_{e_i}(x) = 1$ if $x \in e_i$ and 0 otherwise.

From (17), (18), and (19) it can be seen that the operator $P_N T P_N$ is given by an integral operator with kernel $n_N^k(\cdot, \cdot)$, i.e.,

$$(20) \qquad (P_N T P_N f)(x) = \int_{-A}^{A} n_N^k(x,y) f(y) dy.$$

Then, by using Hölder's inequality, it can be shown from (14) and (20) that

$$(21) \qquad \|T - P_N T P_N\| \le \left(\int_{-A}^{A} \int_{-A}^{A} |k(x,y) - n_N^k(x,y)|^2 dy dx \right)^{\frac{1}{2}}.$$

If the function k is continuous, it is uniformly continuous in the compact interval $[-A, A] \times [-A, A]$. It follows that the functions n_N^k converge uniformly to k. Thus, in the case of continuous k, it follows from the inequality $\| \cdot \|_2 \le 2A \| \cdot \|_\infty$ that the right-hand side of (21) converges to zero as the number of divisions N tends to infinity.

It is a standard result from measure theory that continuous functions on $[-A, A] \times [-A, A]$ constitute a dense subspace of $\mathcal{L}^2([-A, A] \times [-A, A])$ (see, e.g., [11]). Hence, for an arbitrary $k \in \mathcal{L}^2([-A, A] \times [-A, A])$, we can approximate k in the 2-norm by a continuous function g (i.e., $\|k - g\|_2 \to 0$). For any N, the corresponding discrete approximations n_N^k and n_N^g, given by (18) and (19), satisfy

$$\|n_N^k - n_N^g\|_2^2 = \int_{-A}^{A} \int_{-A}^{A} \sum_{i,j} \mathcal{X}_{e_i}(x) \mathcal{X}_{e_j}(y) |k_{i,j} - g_{i,j}|^2 dy dx$$

$$= \sum_{i,j} \left(\frac{2A}{N} \right)^2 \left| \left(\frac{N}{2A} \right)^2 \int_{-A}^{A} \int_{-A}^{A} (k(x,y) - g(x,y)) \mathcal{X}_{e_i}(x) \mathcal{X}_{e_j}(y) dy dx \right|^2$$

$$\le \sum_{i,j} \left(\frac{N}{2A} \right)^2 \|(k-g) \mathcal{X}_{e_i} \mathcal{X}_{e_j}\|_1^2.$$

Now, writing the identity $(k-g)\mathcal{X}_{e_i}\mathcal{X}_{e_j} \equiv [(k-g)\mathcal{X}_{e_i}\mathcal{X}_{e_j}][\mathcal{X}_{e_i}\mathcal{X}_{e_j}]$ and applying Hölder's inequality gives

$$\|n_N^k - n_N^g\|_2^2 \le \sum_{i,j} \left(\frac{N}{2A}\right)^2 \|(k-g)\mathcal{X}_{e_i}\mathcal{X}_{e_j}\|_2^2 \|\mathcal{X}_{e_i}\mathcal{X}_{e_j}\|_2^2$$

(22)
$$= \sum_{i,j} \|(k-g)\mathcal{X}_{e_i}\mathcal{X}_{e_j}\|_2^2 = \|k-g\|_2^2$$

(or we could observe that $k \mapsto n_N^k$ is the orthogonal projection on the span of $\{\mathcal{X}_{e_i}(x)\mathcal{X}_{e_j}(y)\}$, and hence is norm-decreasing).

Since we have already seen that $\|n_N^g - g\|_2 \to 0$ for the continuous function g, and since from (22) we have

(23)
$$\|n_N^k - k\|_2 \le \|n_N^k - n_N^g\|_2 + \|n_N^g - g\|_2 + \|g - k\|_2$$
$$\le 2\|k-g\|_2 + \|n_N^g - g\|_2,$$

we deduce that $\|n_N^k - k\|_2 \to 0$. Therefore, we conclude from (21) that the operator $P_N T P_N$ converges to T in norm.

The transition matrix $\mathbf{Q}$ of the resulting Markov chain is defined by (10) and (12). As explained before, we simplify the computation by taking a piecewise-constant approximation of f_{x_k} (i.e., of the form $\tilde{f}_{x_k} = P_N f_{x_k}$). Then, the elements of the matrix $\mathbf{Q}$ are given by $q_{ij} = \frac{N}{2A} \int_{e_i} \int_{e_j} k(x,y) dy dx$. Therefore, the ith element of the product of the matrix $\mathbf{Q}$ and a vector $\mathbf{p} = \{p_j\}$; $1 \le j \le N$, is given by

(24)
$$(\mathbf{Qp})_i = \frac{N}{2A} \sum_{j=1}^{N} \left[\left(\int_{e_i} \int_{e_j} k(x,y) dy dx \right) p_j \right].$$

From (17) and (24) it can be seen that the matrix $\mathbf{Q}$ (acting on vectors) and the operator $P_N T P_N$ (acting on piecewise-constant functions) are conjugate. To see this, consider a function $f = \sum_i v_i \mathcal{X}_{e_i}$ and a vector $\mathbf{v} = \{v_i\}$; $1 \le i \le N$. Then, by noticing that $\frac{N}{2A} \int_{e_j} f(z) dz = v_j$, it can be seen from (17) and (24) that $P_N T P_N f = \sum_i (\mathbf{Qv})_i \mathcal{X}_{e_i}$. It follows immediately that the eigenvalues of the operator $P_N T P_N$ and those of the matrix $\mathbf{Q}$ are the same, and that the eigenfunctions of $P_N T P_N$ are given by the *piecewise-constant held versions* $f_{\mathbf{vQ}} = \sum_i v_{\mathbf{Q}_i} \mathcal{X}_{e_i}$ of the eigenvectors $\mathbf{v_Q} = \{v_{\mathbf{Q}_i}\}_{i=1}^{N}$ of $\mathbf{Q}$.

It remains to prove that the set of nonzero limit points of the eigenvalues of the operator $P_N T P_N$ is a subset of the set of eigenvalues of the operator T. For this purpose, let's suppose that $\lambda \ne 0$ is a limit point of a sequence of eigenvalues λ_N of $P_N T P_N$; then we have to show that λ is an eigenvalue of T. As we proved before, the operator $P_N T P_N$ converges to T in the natural norm of operators on $\mathcal{L}^2([-A, A])$. Hence, $(\lambda_N I - P_N T P_N)$ converges to $(\lambda I - T)$ in norm. Since the set of invertible operators is open [12, Theorem 10.12], if $(\lambda_N I - P_N T P_N)$ is not invertible (i.e., λ_N belongs to the spectrum of $P_N T P_N$) for all N then $(\lambda I - T)$ is not invertible (i.e., λ belongs to the spectrum of T). Since the operator T is compact [3], and for a compact linear operator every spectral value, with the possible exclusion of zero, is an eigenvalue [5, Theorem 8.4–4], the theorem is proved.   □

4. Example. In order to illustrate the convergence of eigenvalues and eigenvectors we consider the following scalar state space model:
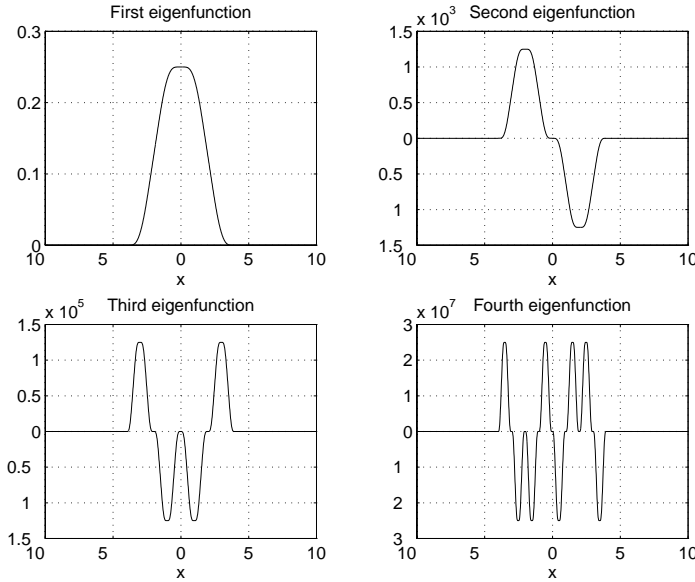
(25)
$$x_{k+1} = 0.5x_k + v_k,$$

FIG. 1. *Eigenfunctions of the operator T corresponding to the four eigenvalues of larger magnitude $(1, 0.5, 0.25,$ and $0.125,$ respectively).*

where $v_k \sim$ Uniform$[-2, 2]$ is an iid sequence, and $x_0 \sim$ Uniform$[-10, 10]$.

Notice that, by (7), (8), and Remark 2.1, the eigenvalues of the operator T defined by (2) for the scalar case are the nonnegative integer powers of $\lambda = 0.5$, and the eigenfunctions of T are the stationary PDF f_{x_∞} and its derivatives. In Figure 1 the stationary PDF f_{x_∞} and its three first derivatives are shown. The eigenfunctions are ordered according to the magnitude of the corresponding eigenvalue (i.e., the first eigenfunction corresponds to the larger eigenvalue, etc.). It is easy to see, for the scalar system (13) and supp $f_v \equiv [-B, B]$, that f_{x_∞} has support equal to $[-|b|B/(1-|a|), |b|B/(1-|a|)]$. For this example, supp $f_{x_\infty} = [-4, 4]$.

Since all distributions in this example have compact support equal or contained in the interval $[-10, 10]$ of the real line, the region in the state space to be discretized is chosen as the interval $\mathcal{X} = [-10, 10]$. This interval is divided into N subsets of size $20/N$ and a Markov chain is defined on the discrete state space $\{e_1, e_2, \ldots, e_N\}$. The transition probability matrix $\mathbf{Q}$ is computed using (10) and (12).

In Figure 2 the eight larger eigenvalues of the matrices $\mathbf{Q}$ corresponding to different values of N (denoted $\lambda_{\mathbf{Q}}(N)$) are shown on a logarithmic scale. Notice that they converge to the eigenvalues corresponding to the continuous state operator T:

$$(26) \qquad \frac{\log(\lambda_{\mathbf{Q}}(N))}{\log(0.5)} \rightarrow \frac{\log(0.5^l)}{\log(0.5)} \equiv l \; ; \;\; \text{for } l = 0, 1, \ldots, 7$$

as $N \rightarrow \infty$. Notice also that the number of points N needed for convergence of the successive eigenvalues is exponentially increasing. The reason for this can be found in Figure 1. The successive eigenfunctions have peaks, the number of which increases proportionally to the powers of two. Therefore, in order to have a faithful representation of them with the eigenvectors of the $\mathbf{Q}$ matrices, the number of turning points of the eigenvectors increases by powers of two. In general, for a particular N,
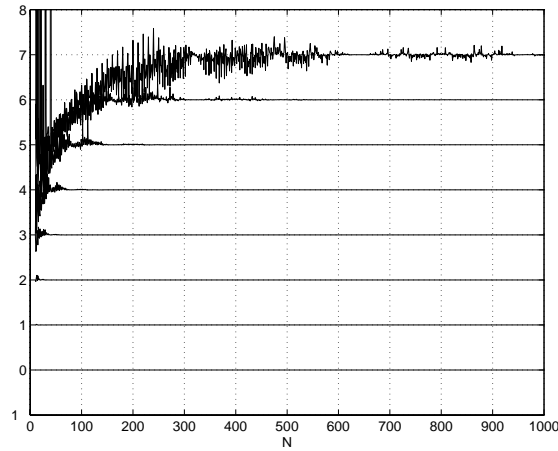
FIG. 2. *Eight larger eigenvalues of matrices* **Q** *on logarithmic scale (*$\log(\lambda_{\mathbf{Q}}(N))/\log(0.5)$*) as a function of the number of subintervals* N.
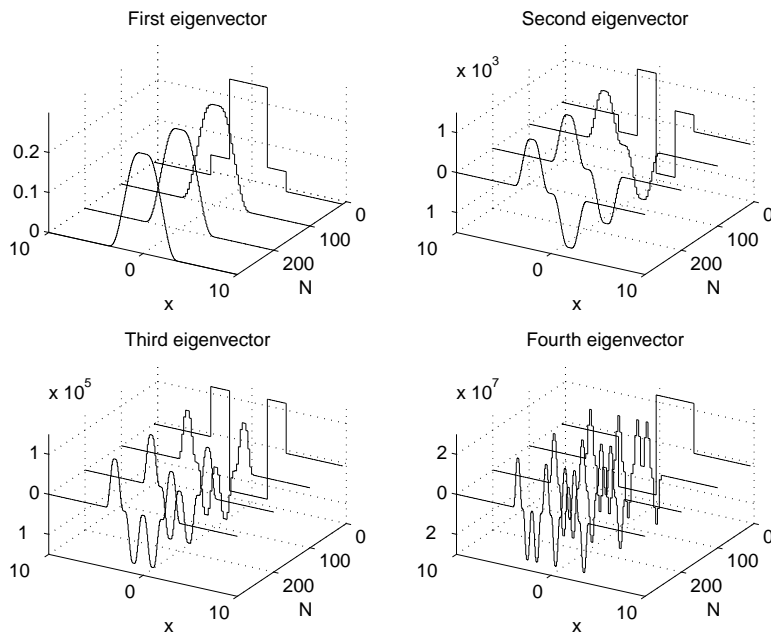

FIG. 3. *Eigenfunctions of* $P_N T P_N$ *corresponding to the four eigenvalues of larger magnitude for* $N = 10$, 100, 200, *and* 300 *subintervals.*

unless an eigenfunction of $P_N T P_N$ approximates one of T, we cannot expect the corresponding eigenvalues to be close approximations.

In Figure 3 the eigenfunctions corresponding to the four eigenvalues of larger modulus of the operators $P_N T P_N$ for $N = 10$, 100, 200, and 300 are shown to illustrate their convergence to the eigenfunctions of T depicted in Figure 1. The eigenfunctions are ordered, as in Figure 1, according to the magnitude of the corresponding eigenvalue.

5. Conclusions. Initial results in the characterization of the class of discrete-state systems amenable to a linear continuous-state approximation have been presented. In particular, we have explored in this paper the properties of discrete-state systems that arise in the inverse process, namely, in the state-discretization of linear continuous-state systems. The method to perform the state-discretization presented consists of a finite rank discrete approximation of the transition integral operator corresponding to the linear continuous-state system. As a result of this state-discretization procedure a Markov chain is obtained. In this work we have shown that the eigenvalues of the resulting Markov chain transition matrices converge to those of the transition operator of the underlying continuous-state linear model. This convergence has been illustrated by a scalar example.

REFERENCES

[1] K. Åström, *Introduction to Stochastic Control Theory*, Academic Press, New York, 1970.
[2] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, London, 1965.
[3] I. Gohberg and S. Goldberg, *Basic Operator Theory*, Birkhäuser, Boston, 1980.
[4] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, New York, Heidelberg, 1992.
[5] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, New York, London, Sydney, 1978.
[6] H. J. Kushner and P. G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, Berlin, New York, Heidelberg, 1992.
[7] D. F. Nicholls and B. G. Quinn, *Random Coefficient Autoregressive Models: An Introduction*, Springer-Verlag, Berlin, New York, Heidelberg, 1982.
[8] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw–Hill, New York, Toronto, London, 1965.
[9] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press, New York, 1981.
[10] J. S. Rosenthal, *Markov chain convergence: From finite to infinite*, Stochastic Process. Appl., 62 (1996), pp. 55–72.
[11] W. Rudin, *Real and Complex Analysis*, McGraw–Hill, London, 1970.
[12] W. Rudin, *Functional Analysis*, McGraw–Hill, New York, Toronto, London, 1973.
[13] P. Terwiesch and M. Agarwal, *A discretized nonlinear state estimator for batch processes*, Comput. Chem. Engrg., 19 (1995), pp. 155–169.
[14] K. B. Wolf, *Integral Transforms in Science and Engineering*, Plenum Press, New York, 1979.