CONTINUUM-STATE HIDDEN MARKOV MODELS WITH DIRICHLET STATE DISTRIBUTIONS

Todd K. Moon and Jacob H. Gunther

Electrical and Computer Engineering Department and Information Dynamics Laboratory
Utah State University
{todd.moon@usu.edu, jake.gunther@usu.edu}

ABSTRACT

In some modeling scenarios, particularly those representing data from natural sources, the discrete states conventionally used in hidden Markov models (HMMs) are at best an approximation, since the discrete states are a modeling artifact. In this paper we present an HMM in which the states take any value in a simplex. The Dirichlet distribution is used to provide a parsimonious representation of the distribution of the states. Conditional state estimates using an extension of the conventional forward/backward method, using Dirichlet distributions to provide a nearly closed-form, but approximate, representation.

1. INTRODUCTION

A conventional hidden Markov model (HMM) is a model in which a sequence of latent (unobserved) state variables $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ form a Markov chain and each element y_t in a sequence of observations $\mathcal{Y}_T = (y_1, y_2, \ldots, y_T)$ is drawn independently of other observations conditional on x_t [1]. State values can be represented as unit S-vectors, with the state $\mathbf{x}_t$ being drawn from the set as

$$\mathbf{x}_t \in \left\{ \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T, \begin{bmatrix} 0 & 1 & \cdots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}^T \right\}. \tag{1}$$

HMMs provide statistical inference procedures in areas such as speech recognition [2], bioinformatics [3], digital communications [4], gesture recognition [5], handwriting recognition [6], human motion recognition [7], etc. In many of these applications, particularly statistical modeling of human or natural activities, states represent a decomposition of the pattern of interest into temporal or spatial components, such as when a phone (speech sound) is represented as having a beginning, middle, and end state. In many such applications, there is a gradual transition between components, so that a "crisp" decomposition into states is a modeling artifact not necessarily present in the physical system being modeled. The greater variability this introduces into the conditional observation distributions has been accommodated, for example, by employing mixture distributions, or by increasing the number of states so that states exist to represent the transitional aspects of the system.

In this paper, the concept of the state of the HMM is generalized so that a state may be any point in the simplex $\Delta_S = \{\mathbf{x} \in \mathbb{R}^S : x_i \geq 0, \sum_{i=1}^S x_i = 1\}$. The vertices of this simplex are the unit vectors indicated in (1) and this generalization of the HMM subsumes the conventional HMM. States are now described not by a pmf, but by a pdf $f(\mathbf{x}_t)$. The distribution $f(\mathbf{x}_t)$ is represented here using a Dirichlet distribution with parameter vector $\boldsymbol{\lambda}_t$, so that $\mathbf{x}_t \sim \mathcal{D}(\mathbf{x}_t; \boldsymbol{\lambda}_t)$. State transitions are described by a conditional pdf $f(\mathbf{x}_{t+1}|\mathbf{x}_t)$, assumed to be a Dirichlet distribution $\mathbf{x}_{t+1}|\mathbf{x}_t \sim \mathcal{D}(\mathbf{x}_{t+1}; \mathbf{l}_{t+1}(\mathbf{x}_t))$. The conditional distribution of the output $f(y_t|\mathbf{x}_t)$ is represented as a superposition (mixture) of pure distributions, with the elements of the conditioning state vector as the mixture weights. Because the state can exist over the continuum of the simplex, the model is referred to as the continuum-HMM, or cHMM. Since the mixture parameters effectively change depending on the state, the cHMM provides for a time-varying, data-dependent mixture.

In section 3, a Kalman filter-like algorithm is developed for estimating the posterior distribution of the state $\mathbf{x}_t$, given observations up to time t. The algorithms for state estimation are formulated on a basis of effectiveness and expediency, sacrificing some accuracy for closed-form representations of the distributions. An approximation to the propagate step has a closed form solution as a Dirichlet distribution, $f(\mathbf{x}_{t+1}|\mathcal{Y}_t) \sim \mathcal{D}(\mathbf{x}_{t+1}|\boldsymbol{\lambda}_{t+t|t})$. The distribution of the update step also has a Dirichlet approximation, $f(\mathbf{x}_{t+1}|\mathcal{Y}_{t+1}) \sim \mathcal{D}(\mathbf{x}_{t+1}|\boldsymbol{\lambda}_{t+1|t+1})$. The Kullback-Leibler (KL) divergence between the true filtered distribution and the Dirichlet approximation is very small.

Since its inception [8] and popularization [1], the HMM has been applied and extended in a variety of ways. In [9], it is recognized that using a finite discrete variable is "unreasonable for most real-world problems." To address this problem, the HMM in [9] is formulated with a countably infinite number of states. However, since the number of states is countable, this differs from our formulation, in which the states exist in the continuum of the simplex. Furthermore, state inference in [9] is by means of Gibbs sampling, which differs significantly from the nearly closed-form solutions presented here. A different generalization is presented in

[10], an HMM is built upon a hierarchical Dirichlet process (HDP). This HMM also has a countably infinite number of states, but is described by only three parameters. As for the first infinite state HMM, inference is still done by means of Gibbs sampling. This model was extended in [11], state persistence modeling was added to the HMM of [10], but differences still remain between this and our cHMM model.

2. CONTINUUM STATE HIDDEN MARKOV MODELS

The state $\mathbf{x}_t$ may any value in the simplex Δ_S. A state $\mathbf{x}_t$ equal to one of the vertices of Δ_S is called a pure state. The distribution of the state $\mathbf{x}_t$ is described using a pdf, which we have chosen to represent as a Dirichlet distribution. A Dirichlet distribution over S variables $\mathbf{x} = (x_1, x_2, \ldots, x_S)$, satisfying the constraints $\sum_{i=1}^S x_i = 1$ and $x_i \geq 0$, is parameterized by a vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_S)$, with $\boldsymbol{\lambda} \in \mathbb{R}_+^S$, that is, each $\lambda_i \geq 0$. The Dirichlet density has the form

$$\begin{aligned} f(\mathbf{x}; \boldsymbol{\lambda}) &= f(x_1, x_2, \ldots, x_S; \lambda_1, \lambda_2, \ldots, \lambda_S) \\ &= \frac{\Gamma(\sum_{i=1}^S \lambda_i)}{\prod_{i=1}^S \Gamma(\lambda_i)} \prod_{i=1}^S x_i^{\lambda_i - 1}, \end{aligned}$$

which is denoted as $\mathbf{x} \sim \mathcal{D}(\mathbf{x}; \boldsymbol{\lambda})$. Where the functional form without the constant is of interest, we will write $f(\mathbf{x}; \boldsymbol{\lambda}) = \mathsf{C} \prod_{i=1}^S x_i^{\lambda_i - 1}$. where, $\mathsf{C} = \Gamma(\sum_i \lambda_i) / \prod_i \Gamma(\lambda_i)$ is a normalizing constant.

We take the conditional distribution of the next state variable as Dirichlet with parameter $\mathbf{l}_{t+1}(\mathbf{x}_t)$: $f(\mathbf{x}_{t+1}|\mathbf{x}_t) \sim \mathcal{D}(\mathbf{x}_{t+1}; \mathbf{l}_{t+1}(\mathbf{x}_t))$. The parameter vector is $\mathbf{l}_{t+1}(\mathbf{x}_t)$, indicating that the Dirichlet parameter vector is (in general) a time-varying function of the conditioning state $\mathbf{x}_t$. The ith element of $\mathbf{l}_{t+1}(\mathbf{x}_t)$ is denoted as $l_{t+1,i}(\mathbf{x}_t)$.

If the state $\mathbf{x}_t$ is a pure state, such as $\mathbf{x}_t = (1, 0, 0)$, then the generalized MM behaves on average like a conventional HMM.

In general, the Dirichlet parameter vector $\mathbf{l}_{t+1}(\mathbf{x}_t)$ could be any function mapping $\Delta_S \to \mathbb{R}_+^S$. However, for reasons of convenience and parsimony, we use a linear function, writing $\mathbf{l}_{t+1}(\mathbf{x}_t) = \sum_{j=1}^S \mathbf{l}_{t+1}(j) x_{t,j}$. The vector $\mathbf{l}_{t+1}(j)$ indicates the parameter of the distribution of the next state when all of the mass of state $\mathbf{x}_t$ is concentrated on state j. Writing the vectors $\mathbf{l}_{t+1}(j)$ as column vectors and stacking these as a matrix (the state transition matrix) $L_{t+1} = \begin{bmatrix} \mathbf{l}_{t+1}(1) & \mathbf{l}_{t+1}(2) & \cdots & \mathbf{l}_{t+1}(S) \end{bmatrix}$, the parameter for the next state conditional distribution is $\mathbf{l}_{t+1}(\mathbf{x}_t) = L_{t+1}\mathbf{x}_t$. With this model, the state transition probability density is $f(\mathbf{x}_{t+1}|\mathbf{x}_t) \sim \mathcal{D}(\mathbf{x}_{t+1}|L_{t+1}\mathbf{x}_t)$, or more explicitly,

$$f(\mathbf{x}_{t+1}|\mathbf{x}_t) = \mathsf{C} \prod_{i=1}^S x_{t+1,i}^{\sum_{j=0}^{S-1} x_{t,j} l_{t+1,i}(j) - 1}$$

The initial state is selected according to a Dirichlet distribution with parameter $\boldsymbol{\lambda}_1 = \Pi$, that is, $\mathbf{x}_1 \sim \mathcal{D}(\mathbf{x}_1; \Pi)$.

A conventional HMM associates with each pure state a distribution governing observations produced in that state. Let $y_t \in \mathbb{R}^d$ denote the observation at time t. If the pure state is i, the distribution of y_t is denoted by $f(y_t|i)$, called the state-constituent distribution. For a state $\mathbf{x}_t \in \Delta_S$, the observation y_t has distribution $f(y_t|\mathbf{x}_t)$. For reasons of convenience and parsimony, we represent $f(y_t|\mathbf{x}_t)$ as a mixture model, in which distributions associated with each pure state are mixed using the state as mixture coefficients,

$$f(y_t|\mathbf{x}_t) = \sum_{i=1}^S x_{t,i} f(y_t|i). \tag{2}$$

Each of the state-constituent distributions $f(y_t|i)$ may take any of the conventional forms for observation densities for HMMs (e.g., Gaussian, mixture model, etc.). The observation model for the generalized state thus subsumes conventional HMMs.

3. STATE ESTIMATION

In this section we develop a method for estimating the distribution of the state of the cHMM from a sequence of observations. The estimator here is a forward-only, Kalman filter-like estimator which provides an estimate of the distribution $f(\mathbf{x}_t|\mathcal{Y}_t)$ of the state $\mathbf{x}_t$ given a sequence of observations $\mathcal{Y}_t = (\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_t)$. A recursive update is provided so that $f(\mathbf{x}_{t+1}|\mathcal{Y}_{t+1})$ can be efficiently computed from $f(\mathbf{x}_t|\mathcal{Y}_t)$. The techniques developed here can be extended re-create the forward backward α/β computations familiar from conventional HMM theory.

The distribution $f(\mathbf{x}_t|\mathcal{Y}_t)$ is assumed to be Dirichlet distributed, $f(\mathbf{x}_t|\mathcal{Y}_t) \sim \mathcal{D}(\mathbf{x}_t; \boldsymbol{\lambda}_{t|t})$, with parameter vector $\boldsymbol{\lambda}_{t|t}$, and at the next time step, the distribution is found to be approximated as $f(\mathbf{x}_{t+1}|\mathcal{Y}_{t+1}) \sim \mathcal{D}(\mathbf{x}_{t+1}; \boldsymbol{\lambda}_{t+1|t+1})$. As is conventional, the development proceeds in two steps, a propagation step and an update step.

The Propagation Step The propagation step computes the distribution of the state at the next time $\mathbf{x}_{t+1}$, given all the observations up to time t, according to $f(\mathbf{x}_{t+1}|\mathcal{Y}_t) = \int_{\Delta_S} f(\mathbf{x}_{t+1}|\mathbf{x}_t) f(\mathbf{x}_t|\mathcal{Y}_t)\, d\mathbf{x}_t$. The update can be represented approximately as a Dirichlet distribution, so that efficient recursive computations are possible. Substituting the definitions for the distributions in, we obtain

$$\begin{aligned} &f(\mathbf{x}_{t+1}|\mathcal{Y}_t) = \\ &= \mathsf{C}_1 \mathsf{C}_2 \frac{1}{\prod_i x_{t+1,i}} \int_{\Delta_S} \prod_{i=1}^S x_{t,i}^{\boldsymbol{\lambda}_{t|t,i} - 1} \Big(\prod_{j=1}^S x_{t+1,j}^{l_{t+1,j}(i)} \Big)^{x_{t,i}} d\mathbf{x}_t. \end{aligned} \tag{3}$$

The integral can be written more abstractly as $\int_{\Delta_S} \prod_{i=1}^S h_i(x_{t,i})\, d\mathbf{x}_t$, where $h_i(x_{t,i}) = x_{t,i}^{\lambda_{t|t,i} - 1} \big(\prod_{j=1}^S x_{t+1,j}^{l_{t+1,j}(i)} \big)^{x_{t,i}}$. Suppressing the time dependence, write this as $h_i(x_i) = x_i^{\lambda_i - 1} e^{-a_i x_i} u(x_i)$, where $u(x)$ is the unit step function (to make this explicitly causal) and where $b_i = \prod_{j=1}^{S} x_{t+1,j}^{l_{t+1,j}(i)}$, and where $a_i = -\ln b_i$. Note that $a_i > 0$.

The integral can be approximately evaluated in closed form by means of the following theorem.

Theorem 1 *Let $h_i(x_i) = x_i^{\lambda_i - 1} e^{-a_i x_i}$. Then*

$$I = \int_{\Delta_S} \prod_{i=1}^{S} h_i(x_i)\, d\mathbf{x}$$

can be approximated as $I \approx e^{-\overline{a}}$, where $\overline{a} = \frac{\sum_{i=1}^{S} a_i \lambda_i}{\sum_{i=1}^{S} \lambda_i}$.

The proof, omitted due to space, explains that the approximation is essentially that of approximating a confluent hypergeometric function by a truncated exponential.

Returning to the update equation (3),

$$f(\mathbf{x}_{t+1}|\mathcal{Y}_t) = \mathsf{C}\frac{1}{\prod_{i=1}^{S} x_{t+1,i}} e^{-\overline{a}} \overset{\approx}{\sim} \mathcal{D}(\mathbf{x}_{t+1}; \boldsymbol{\lambda}_{t+1|t}),$$

where C is a density-normalizing constant and the symbol $\overset{\approx}{\sim}$ means "is approximately distributed as." Thus $f(\mathbf{x}_{t+1}|\mathcal{Y}_t)$ is (approximately) Dirichlet distributed, with parameter vector $\boldsymbol{\lambda}_{t+1|t} = \frac{1}{\sum_{j=1}^{S} \lambda_{t|t,j}} L \boldsymbol{\lambda}_{t|t}$. This establishes the following theorem.

Theorem 2 *Let $f(\mathbf{x}_t|\mathcal{Y}_t) \sim \mathcal{D}(\mathbf{x}_t; \boldsymbol{\lambda}_{t|t})$ and $f(\mathbf{x}_{t+1}|\mathbf{x}_t) \sim \mathcal{D}(\mathbf{x}_{t+1}, \mathbf{l}_{t+1}(\mathbf{x}_t))$. where*

$$\mathbf{l}_{t+1}(\mathbf{x}_t) = \begin{bmatrix}\mathbf{l}_{t+1}(1) & \mathbf{l}_{t+1}(2) & \cdots & \mathbf{l}_{t+1}(S)\end{bmatrix} \mathbf{x}_t \triangleq L_{t+1}\mathbf{x}_t.$$

Then the propagate step is $f(\mathbf{x}_{t+1}|\mathcal{Y}_t) \overset{\approx}{\sim} \mathcal{D}(\mathbf{x}_{t+1}; \boldsymbol{\lambda}_{t+1|t})$, where $\overset{\approx}{\sim}$ means "is approximately distributed as," and where

$$\boldsymbol{\lambda}_{t+1|t} = \frac{1}{\sum_{i=1}^{S} \lambda_{t|t,i}} L_{t+1} \boldsymbol{\lambda}_{t|t}. \tag{4}$$

As an example, consider the case that

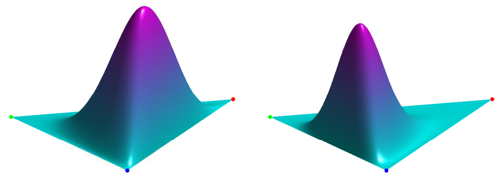
$$\boldsymbol{\lambda}_{t|t} = \begin{bmatrix}3 & 2 & 4\end{bmatrix}^T \qquad \text{and} \qquad L = \begin{bmatrix}5 & 1 & 1\\ 6 & 9 & 1\\ 4 & 2 & 10\end{bmatrix}.$$

The propagated $\boldsymbol{\lambda}_{t+1|t}$, computed using (4) is

$$\boldsymbol{\lambda}_{t+1|t} = \begin{bmatrix}2.33 & 4.44 & 6.22\end{bmatrix}^T.$$

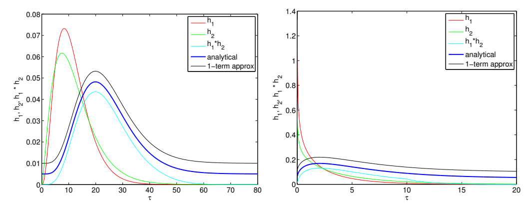
Figure 1(a) shows the distribution $\mathcal{D}(\mathbf{x}; \boldsymbol{\lambda}_{t|t})$, and Figure 1(b) shows the the distribution $\mathcal{D}(\mathbf{x}; \boldsymbol{\lambda}_{t+1|t})$. As a result of the update, the distribution has shifted toward states 2 and 3, as expected.

We present a couple of examples supporting the accuracy of the approximation made in the convolution above. Figure 2(a) shows a function $h_1(\tau) = \tau^{\lambda_1 - 1} e^{-a_1 \tau}$ and a function $h_2(\tau) = \tau^{\lambda_2 - 1} e^{-a_2 \tau}$, the numerical convolution



(a) $\mathcal{D}(\mathbf{x}; \boldsymbol{\lambda}_{t|t})$, $\boldsymbol{\lambda}_{t|t} = (3, 2, 4)$ (b) $\mathcal{D}(\mathbf{x}; \boldsymbol{\lambda}_{t+1|t})$

Fig. 1. Example of propagation



(a) $\lambda_1 = 3.5$, $\lambda_2 = 2.5$, $a_1 = 0.3$, $a_2 = 0.2$ (b) $\lambda_1 = 0.7$, $\lambda_2 = 0.8$, $a_1 = 0.3$, $a_2 = 0.2$

Fig. 2. Comparison of exact and approximate convolutions

$h_1 * h_2$ (cyan), the exact convolution and the approximation of the theorem. In this plot, $\lambda_1 = 3.5$, $\lambda_2 = 2.5$, $a_1 = 0.3$ and $a_2 = 0.2$. Since the differences in the convolutions are nearly imperceptible on the plot, the convolution plots are offset from each other by by 0.05 so that they may be distinguished. To numerical precision, the KL distortion between the analytical convolution and the 1-term approximation is 0 bits.

The Update Step The update step computes $f(\mathbf{x}_{t+1}|\mathcal{Y}_{t+1})$, incorporating the measurement at time $t+1$ into the distribution of the state $\mathbf{x}_{t+1}$. The update step can be written

$$f(\mathbf{x}_{t+1}|\mathcal{Y}_{t+1}) = \frac{f(y_{t+1}|\mathbf{x}_{t+1}, \mathcal{Y}_t)}{f(y_{t+1}|\mathcal{Y}_t)} f(\mathbf{x}_{t+1}|\mathcal{Y}_t).$$

By the assumed Markovian structure, $f(y_{t+1}|\mathbf{x}_{t+1}, \mathcal{Y}_t) = f(y_{t+1}|\mathbf{x}_{t+1})$. Using this and the definition for the output distribution (2),

$$f(\mathbf{x}_{t+1}|\mathcal{Y}_{t+1}) = \mathsf{N}\sum_{i=1}^{S} x_{t+1,i}\, p(y_{t+1}|i) f(\mathbf{x}_{t+1}|\mathcal{Y}_t), \tag{5}$$

where the normalization constant is $\mathsf{N} = \frac{\sum_j \lambda_{t+1|t,j}}{\sum_j f(y_{t+1}|i)\lambda_{t+1|t,j}}$.

Rather than seeking an optimal approximation, we employ the expedience of a simple evaluation technique that attempts to match the measured distribution at $n \geq S$ points.

For the analysis below, it is convenient to use $n = S$ points drawn in from the vertices of the simplex. Let $\mathfrak{c}$ denote the center point of the domain simplex. That is, if $\mathbf{1}$ is the vector of S ones, $\mathfrak{c} = \frac{1}{S}\mathbf{1}$. Let $\mathbf{e}_i$ be the unit vector with 1 in the ith coordinate. Let the vertices be pulled in from the vertices $\mathbf{e}_i$ by a fraction η toward this center point to form n evaluation points $\boldsymbol{\epsilon}_i$ by $\boldsymbol{\epsilon}_i = (1-\eta)\mathbf{e}_i + \eta\mathfrak{c}, i = 1, 2, \ldots, S$. The desired approximate Dirichlet representation is found by evaluating (5) at points $\mathbf{x}_{t+1} = \boldsymbol{\epsilon}_i$ for $i = 1, 2, \ldots, n$, and finding a Dirichlet distribution which matches these evaluated values at these points. Let $g(\mathbf{x}) = f(\mathbf{x}|\mathcal{Y}_{t+1})$. We desire a Dirichlet distribution with parameter vector $\boldsymbol{\lambda}$ (which will represent $\boldsymbol{\lambda}_{t+1|t+1}$) such that

$$g(\mathbf{x})|_{\mathbf{x}=\boldsymbol{\epsilon}_i} = \frac{\Gamma(\sum_{i=1}^S \lambda_i)}{\prod_{i=1}^S \Gamma(\lambda_i)} \prod_{i=1}^S x_i^{\lambda_i - 1}\Bigg|_{\mathbf{x}=\boldsymbol{\epsilon}_i}, \quad i = 1, 2, \ldots, S,$$

or

$$\ln g(\mathbf{x})|_{\mathbf{x}=\boldsymbol{\epsilon}_i} = C + \sum_{j=1}^S (\lambda_j - 1)\ln x_j\Bigg|_{\mathbf{x}=\boldsymbol{\epsilon}_i, i=1,2,\ldots,S} \tag{6}$$

where $C = C(\boldsymbol{\lambda}) = \ln\Gamma(\sum_{i=1}^S \lambda_i) - \sum_{i=1}^S \ln\Gamma(\lambda_i)$. Stacking (6) yields the equations

$$\begin{bmatrix} \ln\epsilon_{1,1} & \ln\epsilon_{1,2} & \cdots & \ln\epsilon_{1,S} \\ \ln\epsilon_{2,1} & \ln\epsilon_{2,2} & \cdots & \ln\epsilon_{2,S} \\ \vdots & & & \\ \ln\epsilon_{S,1} & \ln\epsilon_{S,2} & \cdots & \ln\epsilon_{S,S} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_S \end{bmatrix} = \begin{bmatrix} \ln g(\boldsymbol{\epsilon}_1) + \sum_{j=1}^S \ln x_{1,j} \\ \ln g(\boldsymbol{\epsilon}_2) + \sum_{j=1}^S \ln x_{2,j} \\ \vdots \\ \ln g(\boldsymbol{\epsilon}_S) + \sum_{j=1}^S \ln x_{S,j} \end{bmatrix} - C \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

which can be written as $A\boldsymbol{\lambda} = \mathbf{b} - C\mathbf{1}$. Obviously, if C were known, then the desired parameter $\boldsymbol{\lambda}$ could be readily obtained. But, since C depends on $\boldsymbol{\lambda}$, this is actually a nonlinear equation. We propose an iterative solution. First, define a function $h(C)$ as follows.

---

function $h = h(C)$
  $\boldsymbol{\lambda} = A^\dagger(\mathbf{b} - C\mathbf{1})$
  $h = \ln\Gamma(\sum_{i=1}^S \lambda_i) - \sum_{i=1}^S \ln\Gamma(\lambda_i)$

Here, $A^\dagger$ is the pseudoinverse of A, used if $n > S$ points are used in (6). With $h(C)$ thus defined, an iterative algorithm is proposed. Starting at some initial value of $C = C^{[0]}$, iterates are formed as

$$C^{[k+1]} = h(C^{[k]}). \tag{7}$$

The limit point, when it exists, is denoted as C^∞.

There are generically two fixed points for $h(C)$, as established by the following lemma.

Lemma 1

1. *$h(C)$ is monotonically increasing over its range.*
2. $\lim_{C\to\infty} h'(C) < 1$.
3. *Let C_{min} be the lower limit of the range of $h(C)$. If the elements of $\mathbf{b}$ are not all the same, then $\lim_{C\downarrow C_{min}} h'(C) = \infty$.*
4. *$h(C)$ is a concave function.*

The proof is omitted due to space considerations.

Based on the concavity of the function $h(C)$ there are two fixed points of $h(C)$, one of which is an attractive fixed point, and the other a repelling fixed point.

Let c_1 denote the right (attractive) fixed point of $h(C)$ and c_0 denote the left (repelling) fixed point, with $c_1 > c_0$, and let $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_0$ denote the corresponding λ parameter values for these fixed points. Then it is straightforward to show that $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_1 - (c_1 - c_0)|\sigma_A|\mathbf{1}$. That is, all of the components of $\boldsymbol{\lambda}_0$ are uniformly smaller than corresponding components of $\boldsymbol{\lambda}_1$. Since $\boldsymbol{\lambda}_1$ has larger values than $\boldsymbol{\lambda}_0$, it represents distributions that are more domelike.

For some distributions it may be necessary to solve for the repelling fixed point. This can be done by solving the fixed point directly, for example, numerically solving for a point such that minimizes $(h(C) - C)^2$ using any scalar minimization technique, over a range that includes the left fixed point but not the right fixed point. A convenient point of separation is the point c_- where the function $h(C)$ has unit slope.

4. SUMMARY AND CONCLUSIONS

This paper has presented an HMM providing a continuum of states across a simplex. The number of pure states is a finite number S, so the model is reminiscent of the classical HMM, and in fact (under appropriate parameter settings) subsumes it. The continuity of the states in the new model allows for more nuanced modeling of systems.

Posterior distributions of the state were developed in a Kalman filter-like setting. These distributions were forced (by approximation) to assume the form of Dirichlet distributions. For the examples given the Dirichlet distribution provided an accurate representation of the true distribution. Forcing this form of the distribution was obtained with two approximations. In the first case, the integral over a simplex occuring in the propagation step of a Kalman update is shown to have a closed form representation when a hypergeometric function is approximated as an exponential. The second approximation is obtained by forcing computed values to explicitly approximate a Dirichlet pdf. Finding the parameters of the distribution can be done in nearly closed form, or by means of an iterative scheme.

The results presented here are in many ways preliminary to further analysis of the model. Quantification of the accuracy of the Dirichlet representations imposed on the model has to this point only been achieved numerically, so more analysis of this question is necessary. Convergence of the iterative parameter estimation algorithms remains to be addressed. There remain, too, the broader questions of how this new model will perform in applications.

5. REFERENCES

[1] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, pp. 257–286, February 1989.

[2] M. Gales and S. Young, "The Application of Hidden Markov Models in Speech Recognition," *Foundations and Trends in Signal Processing*, 2007.

[3] P. Baldi and S. Brunak, *Bioinformatics: The machine learning approach.* MIT Press, 2001.

[4] T. K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms.* Wiley, 2005.

[5] H.-K. Lee and J. Kim, "An HMM-based threshold model approach for gesture recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, pp. 961 –973, Oct. 1999.

[6] J. Hu, M. Brown, and W. Turin, "HMM based on-line handwriting recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, pp. 1039 –1045, oct 1996.

[7] X. Zhang and F. Naghdy, "Human Motion Recognition through Fuzzy Hidden Markov Model," in *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, vol. 2, pp. 450 –456, nov. 2005.

[8] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164–171, 1970.

[9] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The Infinite Hidden Markov Model," in *Advances in Neural Information Processing Systems* (T. Dietrich, S. Becker, and Z. Ghahramani, eds.), pp. 577–584, MIT Press, 2002.

[10] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet Processes," *J. Am. Stat. Assoc.*, 2006.

[11] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "An HDP-HMM for systems with state persistence," in *Proceedings of the International Conference on Machine Learning*, July 2008.