Learning and Inference in Parametric Switching Linear Dynamic Systems

Sang Min Oh James M. Rehg Tucker Balch Frank Dellaert

College of Computing, Georgia Institute of Technology
{sangmin, rehg, tucker, dellaert}@cc.gatech.edu

Abstract

We introduce parametric switching linear dynamic systems (P-SLDS) for learning and interpretation of parametrized motion, i.e., motion that exhibits systematic temporal and spatial variations. Our motivating example is the honeybee dance: bees communicate the orientation and distance to food sources through the dance angles and waggle lengths of their stylized dances. Switching linear dynamic systems (SLDS) are a compelling way to model such complex motions. However, SLDS does not provide a means to quantify systematic variations in the motion. Previously, Wilson & Bobick presented parametric HMMs [21], an extension to HMMs with which they successfully interpreted human gestures. Inspired by their work, we similarly extend the standard SLDS model to obtain parametric SLDS. We introduce additional global parameters that represent systematic variations in the motion, and present general expectation-maximization (EM) methods for learning and inference. In the learning phase, P-SLDS learns canonical SLDS model from data. In the inference phase, P-SLDS simultaneously quantifies the global parameters and labels the data. We apply these methods to the automatic interpretation of honey-bee dances, and present both qualitative and quantitative experimental results on actual bee-tracks collected from noisy video data.

1. Introduction

One of the challenging problems in computer vision is the interpretation of video data. Even assuming that targets can be tracked reliably, we encounter the problem of interpreting the resulting trajectories. Manual interpretation, as is often done in domains such as biology, is a time-consuming and error-prone process. Thus, it is desirable to develop methods that automatically interpret the tracks produced by video analysis. In this paper we restrict ourselves to two tasks that are of central importance in video interpretation. The first task is *labeling*, which is to automatically segment the motion according to different behavioral modes. The second task is *quantification*, by which we mean the estimation of global parameters that underly a given motion, e.g. the direction of a pointing gesture.

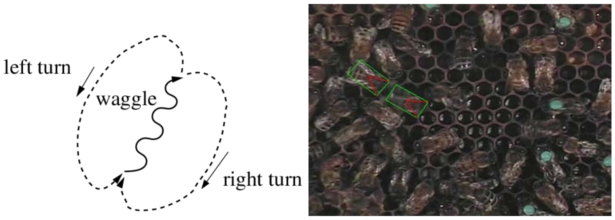


Figure 1: (Left) Three stylized modes of a bee dance. (Right) A vision-based bee tracker.

We take a model-based approach, in which we employ a parameterized computational model of behavior in order to interpret the data. In the case where motions are complex, e.g. they are comprised of sub-behaviors, the model should be expressive enough to capture the interrelationships between the sub-behaviors while it should also be able to model individual sub-behaviors accurately. In this context, a Switching Linear Dynamic System (SLDS) model [14, 15] seems compelling. In an SLDS model, there are multiple linear dynamic systems (LDS) that underly the motion. We can then model the complex behavior of the target by switching within this set of LDSs. In comparison to HMM, SLDS provides the possibility to describe complex temporal patterns concisely and accurately. SLDS models have become increasingly popular in the vision and graphics communities because they provide an intuitive framework for describing the continuous but non-linear dynamics of real-world motion. For example, it has been used for human motion classification [15, 16], and motion synthesis [22].

Nevertheless, standard SLDS models do not provide a means to quantify systematic temporal and spatial variations with respect to a fixed (canonical) underlying behavioral template. The example that motivates us is the honeybee dance. Bees communicate the orientation and distance to food sources through the (spatial) dance angles and (temporal) waggle lengths of their stylized dances which take place in a hive, as illustrated in Fig. 1.

Previously, Wilson and Bobick presented parametric HMMs [21]. In a PHMM, parametric observation models are conditioned on global observation parameters, such that globally parameterized gestures can be recognized. By estimating the global parameters PHMMs have been used to successfully interpret human gestures, showing superior recognition performance in comparison to standard HMMs.

Inspired by this work, we extend the standard SLDS model in a similar manner, resulting in parametric SLDS. We introduce additional global parameters that underly systematic variations of the overall target motion. Moreover, while PHMM only introduced global observation parameters which cause spatial variations, we additionally introduce dynamic parameters which induce temporal variations.

In this paper, we formulate and present expectation-maximization (EM) methods for learning and inference. In the learning phase, P-SLDS learns canonical dynamics from motion data where the individual dynamics may vary due to different underlying global parameters, but we assume these parameters known. In the inference phase, P-SLDS interprets new data, quantifying the global parameters while simultaneously labeling the data.

The remainder of this paper is organized as follows. The standard SLDS model and learning and inference methods are described in Sec. 3. In Sec. 4, we introduce P-SLDS, extending standard SLDS to include global parametric variations. Accordingly, the learning and inference methods for P-SLDS are presented. Lastly, in Sec. 5 we apply P-SLDS to the honeybee dance, and Sec. 6 presents experimental results, comparing the labeling and quantification capabilities of P-SLDS with SLDS.

2. Previous Work

Switching linear dynamic system (SLDS) models have been studied in a variety of research communities ranging from computer vision [3, 14, 12], computer graphics [19, 22], and speech recognition [17] to econometrics [8], machine learning [5, 10, 6, 13], control systems [20] and statistics [18]. SLDS provides natural framework to interpret complex dynamic phenomena. However, exact inference in SLDS is intractable [9]. Thus, there have been research efforts to derive efficient approximation schemes. An early example is GPB2 [1, 3]. More recent examples include a variational approximation [15], expectation propagation [23], sequential Monte Carlo methods [4], Gibbs sampling [17] and Data-Driven MCMC [13]. The SLDS learning problem is studied from the control systems perspective in [20].

In the computer vision community, Pavlović et al. [14] applied SLDS to human motion analysis. They introduced both an approximate Viterbi method and a variational approximation method and compared these methods with GPB2 and HMMs. In related work, North et al. explored the use of switching component models to automatically classify the motion patterns of rigid objects or human body motions [12]. Howard and Jebara have proposed a tree-structured extension of SLDS and applied it to the classification of football plays [6]. The graphics community has modeled video and motion capture data with a set of switching components to obtain dynamic textures [19] and motion textures [22]. SLDS models have also been used to quantify the naturalness of human motion [16].

Parametric HMM (PHMM) models were introduced and applied to human gesture recognition in [21]. In related work, Brand and Hertzmann [2] introduced style machines, a kind of parametric HMM in which the observation model is parameterized by style variables. Style machines were used to learn a set of common dance styles from a varying set of dance sequences and to synthesize novel motions.

3. SLDS Background

A switching linear dynamic system (SLDS) model describes the dynamics of a complex physical process by switching between a set of linear dynamic systems (LDS). Each LDS describes a local dynamic process which is assumed to be linear and Gaussian, and transitions between LDS models are described by a Markov transition matrix.

3.1. Linear Dynamic Systems

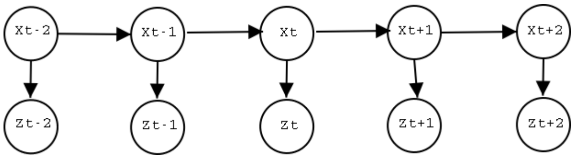


Figure 2: Linear dynamic system (LDS)

We first review linear dynamic systems. An LDS is a time-series state-space model that consists of a linear dynamics model and a linear observation model. The representation of an LDS as a graphical model is shown in Fig. 2. The top Markov chain represents the state evolution of the continuous hidden states x_t. The prior density p_1 on the initial state x_1 is assumed to be normal with mean μ_1 and covariance Σ_1: $x_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$.

The state x_t is obtained by the product of the state transition matrix F and the previous state x_{t-1} corrupted by the additive white noise w_t, zero-mean and normally distributed with covariance matrix Q:

$$x_t = Fx_{t-1} + w_t \text{ where } w_t \sim \mathcal{N}(0, Q) \tag{1}$$

In addition, the measurement z_t is generated from the current state x_t through the observation matrix H, and corrupted by white observation noise v_t:

$$z_t = Hx_t + v_t \text{ where } v_t \sim \mathcal{N}(0, V) \tag{2}$$

Thus, an LDS model M is defined by the tuple $M \triangleq \{(\mu_1, \Sigma_1), (F, Q), (H, V)\}$. Exact inference in an LDS can be done efficiently through Kalman-smoothing [1].

3.2. Switching Linear Dynamic Systems

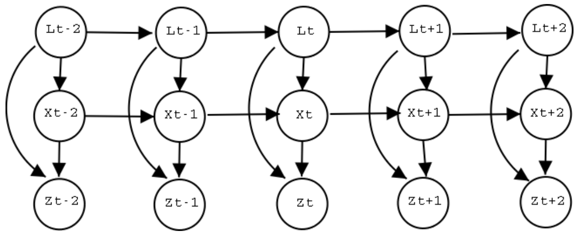


Figure 3: Switching Linear Dynamic System (SLDS)

A *switching* LDS is a natural extension of an LDS, where we assume the existence of n distinct LDS models $M \triangleq \{M_l | 1 \leq l \leq n\}$, where each model M_l is defined by the parameters described in Sec. 3.1. The graphical model corresponding to an SLDS is shown in Fig. 3. The middle chain, representing the hidden state sequence $X \triangleq \{x_t | 1 \leq t \leq T\}$, together with the observations $Z \triangleq \{z_t | 1 \leq t \leq T\}$ at the bottom, is identical to the LDS in Fig. 2. However, we now have an additional discrete Markov chain $L \triangleq \{l_t | 1 \leq t \leq T\}$ that determines which of the n models M_l is being used at every time-step. We call $l_t \in M$ the *label* at time t and L a *label sequence*.

In addition to the set of LDS models M, we specify two additional parameters: a multinomial distribution $\pi \triangleq P(l_1)$ over the initial label l_1 and an $n \times n$ transition matrix T that defines the switching behavior between the n distinct LDS models, i.e. $T_{ij} \triangleq P(l_j | l_i)$. In summary, an SLDS model is completely defined by the tuple $\Theta \triangleq \left\{\pi, T, M \triangleq \{M_l | 1 \leq l \leq n\}\right\}$, which we refer to as the *canonical parameters*.

3.3. Learning an SLDS via EM

The expectation-maximization (EM) algorithm [11] can be used to learn the maximum-likelihood (ML) parameters $\hat{\Theta}$:

$$\hat{\Theta} \triangleq \underset{\Theta}{\operatorname{argmax}} \;\; P(Z|\Theta)$$

The hidden variables in EM are the label sequence L and the state sequence X, i.e. the top and the middle chain in Fig. 3. Given the observation data Z, EM iterates between the following two steps:

- E-step: obtain the posterior distribution

  $$f^i(L, X) \triangleq P(L, X | Z, \Theta^i) \tag{3}$$

  over the hidden variables L and X, using the current estimate for the SLDS parameters Θ^i.

- M-step: maximize the expected log-likelihoods

  $$\Theta^{i+1} \leftarrow \underset{\Theta}{\operatorname{argmax}} \;\; \langle \log P(L, X, Z | \Theta) \rangle_{f^i(L,X)} \tag{4}$$

  Above, $\langle \cdot \rangle_p$ denotes the expectation of a function $(\cdot)$ under a distribution p. As discussed in Sec. 2, the exact E-step in Eq. 3 is intractable and therefore approximate inference methods must be employed.

4. Parametric SLDS

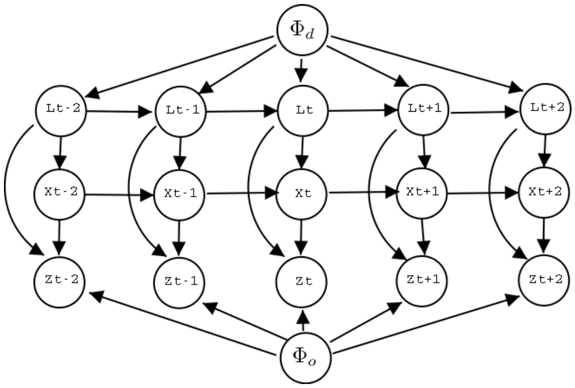


Figure 4: Parametric SLDS (P-SLDS)

We now develop the parametric SLDS (P-SLDS) model, where the discrete state transition probabilities and output probabilities are parameterized by a set of *global* parameters $\Phi = \{\Phi_d, \Phi_o\}$. The parameters Φ are global in that they affect the entire sequence in a systematic way. The graphical model for P-SLDS is shown in Fig. 4. Note that there are two classes of global parameters: the dynamics parameters Φ_d and the observation parameters Φ_o.

The *dynamics parameters* Φ_d represent the factors that cause *temporal* variations. The different values of the dynamics parameters Φ_d result in different switching behaviors between the behavioral modes. For example, in the bee-dance, a food source that is far away leads a dancer bee to stay in each dance regime longer in order to make a larger dance. In contrast, the *observation parameters* Φ_o represent factors that cause *spatial* variations. A good example is a pointing gesture, where the overall arm motion changes as a function of the pointing direction.

The canonical parameters Θ represent the common underlying behavioral templates. Note that the Θ parameters are embedded in the conditional dependency arcs in Fig. 4. In the bee dancing example, the canonical parameters may describe the prototyped stylized bee dance. However, the dynamics of individual bee dances vary systematically from the prototype dance due to the changing food source locations. These are the variations that are represented by the global parameters.

Note that the discrete state transitions in the top chain of Fig. 4 are instantiated by Θ and Φ_d, and the observation model at the bottom is instantiated by Θ and Φ_d, while the continuous state transitions in the middle chain are instantiated solely by the canonical parameters Θ. In other words, the dynamics parameters Φ_d vary the prototyped switching behaviors and the observation parameters Φ_o vary the prototyped observation model. For a given data set, estimation of the global parameters effectively identifies the discrepancies between the observations and dynamics of the data and the behavioral template which is defined by the canonical parameters.

The graphical model of P-SLDS necessitates parameterized versions of an initial state distribution $P(l_1|\Theta, \Phi_d)$, a discrete state transition table $P(l_t|l_{t-1}, \Theta, \Phi_d)$ and an observation model $P(z_t|l_t, x_t, \Theta, \Phi_o)$. In general, there are three possibilities for the nature of the parameterization: (a) the PDF is a linear function of the global parameters Φ, (b) the PDF is a non-linear function of Φ, and (c) no functional form for the PDF is available. In the latter case, a neural network may be used, as discussed in [21]. In our development of learning and inference methods for P-SLDS in Sec. 4.1 and 4.2, we will assume that functional forms are available.

4.1. Learning in P-SLDS

Learning in P-SLDS involves estimating the P-SLDS parameters Θ, given the data $\bar{D} \triangleq \left\{\bar{\Phi} = \{\bar{\Phi}_d, \bar{\Phi}_o\}, \bar{L}, \bar{Z}\right\}$. The elements of $\bar{D}$ are: a set of global parameters $\bar{\Phi} = \{\bar{\Phi}_d, \bar{\Phi}_o\}$, a label sequence $\bar{L}$, and the observations $\bar{Z}$. The upper bars indicate that the values are known. We employ EM to find an ML estimate of the canonical parameters $\hat{\Theta}$ in the presence of the hidden continuous states X. The steps are outlined in Algorithm 1.

The E-step in Eq. 5 is equivalent to inference in an LDS model. In more detail, as the global parameters $\bar{\Phi}$, the current P-SLDS parameters Θ^i, the label sequence $\bar{L}$, and the observations $\bar{Z}$ are all known, the inference over the continuous hidden states X in the E-step can be performed by Kalman smoothing. Given the posterior distribution $f_L^i(X)$ from Eq. 5, the M-step produces an update for the parameters Θ^{i+1}.

- **E-step 1:** obtain the posterior distribution

  $$f_L^i(X) \triangleq P(X|\Theta^i, \bar{D}) \tag{5}$$

  over the hidden state sequence X, based on the current estimate of the canonical parameters Θ^i.

- **M-step 1**: maximize the expected log-likelihood :

  $$\Theta^{i+1} \leftarrow \underset{\Theta}{\text{argmax}} \left\langle \log P(\bar{L}, X, \bar{Z}|\Theta, \bar{\Phi}) \right\rangle_{f_L^i(X)} \tag{6}$$

Algorithm 1: EM1 for Learning in P-SLDS

In the case where the parameterized dependencies are linear functions of the global parameters Φ, the M-step in Eq. 6 can be solved analytically. However, in the case where the parametric dependencies are non-linear, an exact M-step is usually infeasible and must be obtained by alternative optimization methods, such as conjugate gradient or Levenberg-Marquardt.

4.2. Inference in P-SLDS

Given the observations $\bar{Z}$, we can use the learned P-SLDS canonical parameters Θ to estimate the global parameters Φ and infer the label sequence L. Note that the canonical parameters Θ are fixed after they are learned from the training dataset $\bar{D}$. The addition of the global parameters makes it possible to adapt the model to the specific characteristics of each observation sequence, resulting in improved estimation of the hidden labels.

The EM method for estimating the optimal global parameters $\hat{\Phi}$ is shown in Algorithm 2. Note that we use EM1 to learn the canonical model parameters Θ and EM2 to estimate the global parameters Φ and the hidden labels L. We now describe the details of EM2. In the following sections, we use $\mathcal{LLH}$ to denote the log-likelihood.

4.2.1. E-step 2

Approximate inference methods are required since the exact E-step in Eq. 8 is known to be intractable [9]. We adopt the approximate Viterbi method described in [15], as it is simple and fast, and is usually comparable to other methods in the quality of its estimates. At every i^{th} EM iteration, the joint posterior over the hidden variables L and X is approximated by a peaked posterior over X with the obtained

- **E-step 2 :** obtain the posterior distribution :

$$f_I^i(L,X) \triangleq P(L,X|\bar{Z},\Theta,\Phi^i) \tag{8}$$

  over the hidden label sequence L and the state sequence X, using the current estimate for the global parameters Φ^i.

- **M-step 2 :** maximize the expected log-likelihood:

$$\Phi^{i+1} \leftarrow \underset{\Phi}{\operatorname{argmax}} \ \left\langle \log P(L,X,\bar{Z}|\Theta,\Phi) \right\rangle_{f_I^i(L,X)} \tag{9}$$

Algorithm 2: EM2 for Inference in P-SLDS

pseudo-optimal label sequence $\hat{L}^i$:

$$\begin{aligned} P(L,X|\bar{Z},\Phi^i) &= P(X|L,\bar{Z},\Phi^i)P(L|\bar{Z},\Phi^i) \\ &\approx P(X|\hat{L}^i,\bar{Z},\Phi^i)\delta(\hat{L}^i) \end{aligned} \tag{7}$$

$$f_I^i(X) \triangleq P(X|\hat{L}^i,\bar{Z},\Phi^i)\delta(\hat{L}^i)$$

Note that the implicit conditional dependence on the fixed canonical parameters Θ is omitted for clarity.

4.2.2. M-step 2

Using the approximate posterior $f_I^i(X)$ obtained in Eq. 7, the expected complete log-likelihood ($\mathcal{LLH}$) in Eq. 9 is approximated as:

$$\begin{aligned} \mathcal{L}^i(\Phi) &\triangleq \sum_L \int_X \log P(L,X,\bar{Z}|\Phi)P(L,X|\bar{Z},\Phi^i) \\ &\approx \int_X \log P(\hat{L}^i,X,\bar{Z}|\Phi)f_I^i(X) \end{aligned} \tag{10}$$

Using the chain rule, this factors as:

$$P(\hat{L}^i,X,\bar{Z}|\Phi) = P(\hat{L}^i|\Phi_d)P(X,\bar{Z}|\hat{L}^i,\Phi_o) \tag{11}$$

Note that we now only condition on relevant global parameters, e.g. the label sequence $\hat{L}^i$ is only conditioned on Φ_d. Substituting (11) into the expected $\mathcal{LLH}$ $\mathcal{L}^i(\Phi)$ (10), we obtain a more succinct form of $\mathcal{L}^i(\Phi)$ in which the term $\log P(\hat{L}^i|\Phi_d)$ is moved outside the integral:

$$\begin{aligned} \mathcal{L}^i(\Phi) &= \log P(\hat{L}^i|\Phi_d) + \int_X \log P(X,\bar{Z}|\hat{L}^i,\Phi_o)f_I^i(X) \\ &= \mathcal{L}^i(\Phi_d) + \mathcal{L}^i(\Phi_o) \end{aligned} \tag{12}$$

Here we introduced two convenience terms, the *dynamic log-likelihood* $\mathcal{L}(\Phi_d)$ and the *observation log-likelihood* $\mathcal{L}(\Phi_o)$:

$$\mathcal{L}^i(\Phi_d) \triangleq \log P(\hat{L}^i|\Phi_d) \tag{13}$$

$$\mathcal{L}^i(\Phi_o) \triangleq \int_X \log P(X,\bar{Z}|\hat{L}^i,\Phi_o)f_I^i(X) \tag{14}$$

In Eq. 12, we can observe that the total expected $\mathcal{LLH}$ $\mathcal{L}^i(\Phi)$ is maximized by *independently* updating the global observation parameters Φ_o and dynamic parameters Φ_d, i.e. we obtain the updated global parameters Φ_d^{i+1} and Φ_o^{i+1} by maximizing the dynamic $\mathcal{LLH}$ $\mathcal{L}^i(\Phi_d)$ and the observation $\mathcal{LLH}$ $\mathcal{L}^i(\Phi_o)$ respectively.

Now we can further factorize the dynamic $\mathcal{LLH}$ $\mathcal{L}^i(\Phi_d)$ from Eq. 13 and the observation $\mathcal{LLH}$ $\mathcal{L}^i(\Phi_o)$ from Eq. 14, obtaining:

$$\mathcal{L}^i(\Phi_d) = \log P(\hat{l}_1^i|\Phi_d) + \log \sum_{t=2}^{|Z|} P(\hat{l}_t^i|\hat{l}_{t-1}^i,\Phi_d) \tag{15}$$

$$\begin{aligned} \mathcal{L}^i(\Phi_o) &= \int_X \log \left\{ P(\bar{Z}|X,\hat{L}^i,\Phi_o)P(X|\hat{L}^i) \right\} f_I^i(X) \\ &\equiv \int_X \log P(\bar{Z}|X,\hat{L}^i,\Phi_o)f_I^i(X) \\ &= \sum_{t=1}^{|Z|} \int_{x_t} \log P(\bar{z}_t|x_t,\hat{l}_t^i,\Phi_o)f_I^i(x_t), \end{aligned} \tag{16}$$

where the term $f_I^i(x_t)$ denotes the marginal on x_t from the full posterior $f_I^i(X)$, i.e. $f_I^i(x_t) \triangleq \int_{X/x_t} f_I^i(X)$.

The details of the M-step will depend upon the application domain. In the case where the parametric forms are linear in the global parameters Φ, the M-step is analytically feasible. Otherwise, alternative optimization methods can be used to maximize the non-linear $\mathcal{LLH}$ function, as described in Section 4.1.

5. Bee Dance Modeling

We have applied the P-SLDS model to the honeybee dance, with the aim of providing field biologists with a new tool for the quantitative study of insect behavior. Measurements of real-world dancer bee tracks are obtained from video data using a previously developed tracker [7], see Fig. 1. Given the stylized nature of the bee dance, we adopt an approach which decomposes the dance into three different regimes : "turn left", "turn right" and "waggle", illustrated in Fig. 1.

The bee dance is parameterized by both classes of global parameters. The global dynamics parameter set $\Phi_d \triangleq \{\Phi_{d,i}|1 \le i \le n\}$ is chosen to be correlated with the average length of each dance regimes where $n = 3$. The global observation parameter Φ_o is set to be the angle of orientation of the bee dance.

The specific form of the parameterized discrete state transition table, $T(\Phi_d) \triangleq P(l_t|l_{t-1},\Theta,\Phi_d)$, is given by

$$T(\Phi_d)_{ij} = \begin{cases} 1-\Phi_{d,i} & \text{if } l_i = l_j \\ \frac{\Phi_{d,i}}{n-1} & \text{otherwise} \end{cases} \tag{17}$$

In Eq. 17, row i of the Markov transition matrix $T(\Phi_d)$ depends on the global dynamics parameter $\Phi_{d,i}$ where $T(\Phi_d)_{i,j} \triangleq P(l_j|l_i, \Phi_{d,i})$. Here $\Phi_{d,i}$ denotes the probability of a transition out of state i. The M-step update for each $\Phi_{d,i}$ can be obtained by differentiating $\mathcal{LLH}$ in Eq. 15 and normalizing to obtain

$$\Phi_{d,i}^{new} \quad \leftarrow \quad \frac{C_2(i)}{C_1(i) + C_2(i)}. \tag{18}$$

The term $C_1(i)$ above denotes the self-transition counts from the state i to itself in the current Viterbi label sequence $\hat{L}$, and whereas $C_2(i)$ denotes the transition counts from state i to all others.

The parameterized observation model $P(z_t|l_t, x_t, \Phi_o)$ is defined by

$$z_t \quad \sim \quad \mathcal{N}(R(\Phi_o)H_{\hat{l_t}}x_t, V_{\hat{l_t}}), \tag{19}$$

where $R(\Phi_o)$ is the rotation matrix, and $H_{\hat{l_t}}$ and $V_{\hat{l_t}}$ denote the observation parameters of the $\hat{l_t}$th component LDS, corresponding to label $\hat{l_t}$ of the Viterbi sequence $\hat{L}$. We also define $\alpha_t(\Phi_o)$ to be the projected-then-rotated vector of the corresponding state x_t:

$$\alpha_t(\Phi_o) \quad \triangleq \quad R(\Phi_o)H_{l_t}x_t \tag{20}$$

Combining Eq. 19 and 20, we obtain the observation $\mathcal{LLH}$ $\mathcal{L}^i(\Phi_o) \equiv$

$$-\sum_{t=1}^{|Z|} \left\langle [z_t - \alpha_t(\Phi_o)]^T \, V_{\hat{l_t}}^{-1} \, [z_t - \alpha_t(\Phi_o)] \right\rangle_{f_I^i(x_t)} \tag{21}$$

where we have omitted redundant constant terms. Intuitively, the goal in optimizing Eq. 21 is to find an updated dance angle Φ_o^{i+1} which minimizes the sum of the expected Mahalanobis distances between the observations z_t and the projected-then-rotated states $\alpha_t(\Phi_o)$. Since nonlinearities are involved as a consequence of the rotation, there is no analytic solution to the maximization problem in Eq. 21. Therefore, we perform 1D gradient ascent to obtain a numerical solution.

6. Experimental Results

Our experimental results show that P-SLDS provides reliable global parameter estimation capabilities, along with improved recognition performance in comparison to standard SLDS models. Six dancer bee tracks obtained from video are shown in Fig. 5. Fig. 1 displays a video frame from the automatic vision-based tracker [7] which was used to obtain the tracks in Fig. 5. The rectangular bounding boxes denote tracked bees.

We performed experiments using 6 video sequences with lengths 1058, 1125, 1054, 757, 609 and 814 frames. The tracker produces a time-series sequence of vectors $z_t = [x_t, y_t, \cos(\theta_t), \sin(\theta_t)]^T$ where x_t,y_t and θ_t denote the 2D coordinates and the heading angle at time t, respectively. Note that the observed heading angle θ_t differs from the global dance angle Φ_o. Note from Fig. 5 that the tracks are noisy and much more irregular than the stylized dance prototype illustrated in Fig. 1. The red, green and blue colors represent right-turn, waggle and left-turn phases. The ground-truth labels are marked manually for the comparison and learning purposes. The dimensionality of the continuous hidden states was four.

Given the relative difficulty of obtaining this data, which has to be labeled manually to allow for a ground-truth comparison, we adopted a leave-one-out strategy. The parameters are learned from five out of six datasets, and the learned model is applied to the left-out dataset to perform the angle/average waggle length (AWL) quantification and simultaneous labeling. Six experiments are carried out using both P-SLDS and the original SLDS. The P-SLDS estimates of angle/AWL are directly obtained from the results of global parameter quantification. On the other hand, the SLDS estimates are obtained for comparison purposes by averaging the transition numbers and averaging the heading angle over the inferred waggle segments.

6.1. Qualitative Results

The experimental results show the superior recognition capabilities of the proposed P-SLDS model over the original SLDS model. The label inference results for sequences 1, 2, and 6 are shown in Fig. 6. In each figure, the x-axis is the time in frames and the color encodes the corresponding label for each video frame. The results for the other three sequences are comparable to those in Fig. 6.

The superior recognition abilities of P-SLDS can be observed from the presented results. The P-SLDS results match the ground truth more closely than the SLDS results. In particular, Sequence 6 (see Fig. 5(6)) is very noisy. It has very short waggle (green) phases and its dance angle diverges far from the other sequences. The result is a challenging inference problem. Nevertheless, P-SLDS detected several waggle phases correctly while SLDS detected none (see Fig. 6c). This is possible because P-SLDS uses the additional global parameter information to robustly discern the subtle differences that characterize the dance regimes.

6.2. Quantitative Results

Quantitative results for the estimation of the angle and average waggle length of the dances show the robust global parameter estimation capabilities of P-SLDS. Table 1 shows

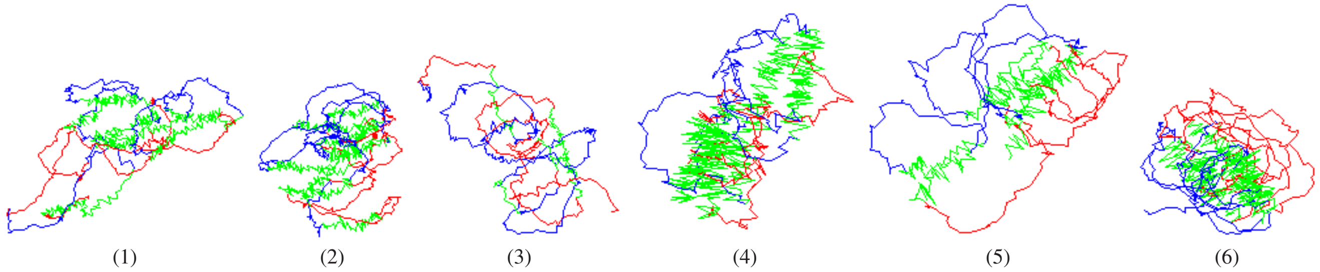
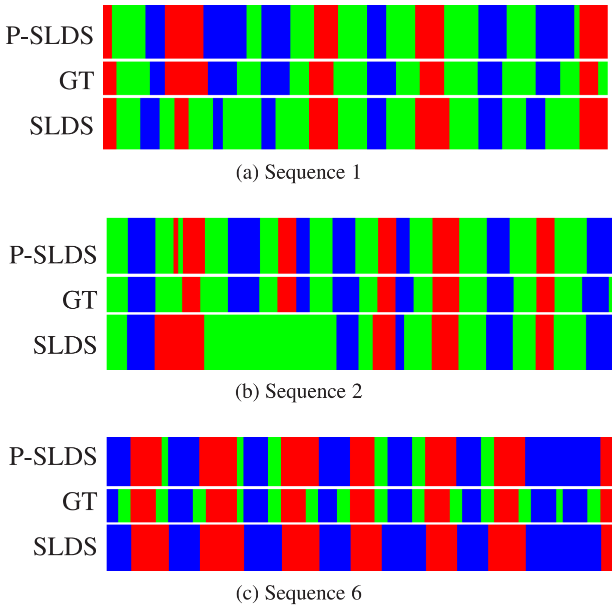


Figure 5: Bee dance sequences used in the experiments. Each dance trajectory is the output of a vision-based tracker. Tables 1 and 2 give the global motion parameters for each of the numbered sequences.



(a) Sequence 1

(b) Sequence 2

(c) Sequence 6

Figure 6: Label inference results. Estimates from P-SLDS and SLDS models are compared to manually-obtained ground truth (GT) labels. Key: Waggle, Right, Left.

Sequence	1	2	3	4	5	6
P-SLDS	0.13	0.13	0.02	0.10	0.03	0.06
SLDS	0.24	0.04	0.71	0.12	0.12	-
GT	-0.30	-0.25	1.13	-1.30	0.80	-2.08

Table 1: Errors in the global rotation angle estimates from P-SLDS and SLDS in radians. Last row contains the ground truth rotation angles. Sequence numbers refer to Fig. 5.

Sequence	1	2	3	4	5	6
P-SLDS	-6.3	-3.0	+6.6	-1.1	-6.9	-2.7
SLDS	+0.1	+35.9	+21.9	+2.2	-4.6	-
GT	57.9	51.2	21.4	41.1	32.6	19.4

Table 2: Errors in the Average Waggle Length (AWL) estimates for P-SLDS and SLDS in frames. Last row contains the ground truth AWL. Sequence numbers refer to Fig. 5.

the errors in angle between the ground truth and the P-SLDS/SLDS estimates. The angle errors of P-SLDS are acceptable, ranging as they do from 0.03 to 0.13 radians. In contrast, it can be observed that the SLDS angle estimates, which are heuristically obtained, are inconsistent. For sequence 6, no angle estimate is available as no waggle segment was detected. The SLDS errors range from 0.04 to 0.71 radians.

Similarly, the quantitative results for average waggle length (AWL) estimation show that P-SLDS can also robustly quantify the global dynamics parameters. AWL is an indicator of the distance to the food source from the hive. Reliable estimates of AWL are of value to insect biologists. Table 2 shows the errors in the P-SLDS and SLDS estimates, along with the ground truth. The P-SLDS estimates were obtained from the global dynamics parameters, while the SLDS estimates were obtained by averaging over the estimated waggle segments.

The results demonstrate that the P-SLDS estimates match the ground-truth closely. The absolute errors of P-SLDS range from 1.1 to 6.9 frames. In contrast, it is observed that the SLDS estimates are inaccurate. More specifically, no estimate from SLDS is available for sequence 6 as no waggle segment is recognized. In addition, the errors on sequences 2 and 3 are 35.9 and 21.9 frames respectively, which indicate that the SLDS estimates are untrustworthy.

7. Conclusions and Future Work

We have introduced the parametric SLDS (P-SLDS) model, a novel parametric extension of SLDS, and presented EM algorithms for inference and learning. The addition of global parameters allows the P-SLDS model to explain systematic variations in the dynamics and observation properties of input data. This has three main benefits: First, we can learn the canonical dynamics and observation models from training sequences whose individual characteristics may vary due to global parameter changes.

Second, we can use inference within the P-SLDS model to estimate the global parameters robustly, leading to improved performance in estimating hidden labels based on the learned models. Third, in applications such as biotracking, the global parameters may have intrinsic meaning. This is the case for the bee dance, where the global parameters encode information about food source location. Our experimental results with real-world bee dance data demonstrate the benefits of the P-SLDS framework.

One avenue for future work is to incorporate additional global parameters, such as affine transformations, into the P-SLDS model. We also plan to explore the use of the P-SLDS model in other application domains, such as vision-based tracking and time-series visualization.

Acknowledgements

The authors would like to thank Zia Khan and Grant Schindler for providing the bee tracking data. This material is based upon work which was supported in part by the National Science Foundation under NSF Awards IIS-0219850 and IIS-0133779. The first author is supported by a fellowship from the Samsung Lee Kun Hee Scholarship Foundation.

References

[1] Y. Bar-Shalom and X. Li. *Estimation and Tracking: principles, techniques and software*. Artech House, Boston, London, 1993.

[2] M. Brand and A. Hertzmann. Style machines. In *Siggraph 2000, Computer Graphics Proceedings*, pages 183–192, 2000.

[3] C. Bregler. Learning and recognizing human dynamics in video sequences. In *Proc. CVPR*, 1997.

[4] A. Doucet, N. J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Trans. Signal Processing*, 49(3), 2001.

[5] Z. Ghahramani and G. E. Hinton. Variational learning for switching state-space models. *Neural Computation*, 12(4):963–996, 1998.

[6] A. Howard and T. Jebara. Dynamical systems trees. In *Conf. on Uncertainty in Artificial Intelligence*, pages 260–267, Banff, Canada, 2004.

[7] Z. Khan, T. Balch, and F. Dellaert. A Rao-Blackwellized particle filter for EigenTracking. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[8] C.-J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60, 1994.

[9] U. Lerner and R. Parr. Inference in hybrid networks: Theoretical limits and practical algorithms. In *Proc. 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01)*, pages 310–318, Seattle, WA, 2001.

[10] U. Lerner, R. Parr, D. Koller, and G. Biswas. Bayesian fault detection and diagnosis in dynamic systems. In *Proc. AAAI*, Austin, TX, 2000.

[11] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 1997.

[12] B. North, A. Blake, M. Isard, and J. Rottscher. Learning and classification of complex dynamics. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(9):1016–1034, 2000.

[13] S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert. Data-driven MCMC for learning and inference in switching linear dynamic systems. In *AAAI Nat. Conf. on Artificial Intelligence*, 2005.

[14] V. Pavlović, J. M. Rehg, T.-J. Cham, and K. Murphy. A dynamic Bayesian network approach to figure tracking using learned dynamic models. In *Intl. Conf. on Computer Vision (ICCV)*, volume 1, pages 94–101, 1999.

[15] V. Pavlović, J. M. Rehg, and J. MacCormick. Learning switching linear models of human motion. In *Advances in Neural Information Processing Systems (NIPS)*, pages 981–987, 2000.

[16] L. Ren, A. Patrick, A. Efros, J. Hodgins, and J. M. Rehg. A data-driven approach to quantifying natural human motion. *ACM Trans. on Graphics, Special Issue: Proc. of 2005 SIGGRAPH Conf.*, 2005. Accepted for publication.

[17] A.-V. Rosti and M. Gales. Rao-blackwellised Gibbs sampling for switching linear dynamical systems. In *Intl. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, volume 1, pages 809–812, 2004.

[18] R. Shumway and D. Stoffer. Dynamic linear models with switching. *Journal of the American Statistical Association*, 86:763–769, 1992.

[19] S. Soatto, G. Doretto, and Y. Wu. Dynamic Textures. In *Intl. Conf. on Computer Vision (ICCV)*, pages 439–446, 2001.

[20] R. Vidal, A. Chiuso, and S. Soatto. Observability and identifiability of jump linear systems. In *Proc. IEEE Conf. on Decision and Control (CDC 02)*, 2002.

[21] A. D. Wilson and A. F. Bobick. Parametric hidden Markov models for gesture recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 21(9):884–900, 1999.

[22] Y.Li, T.Wang, and H.-Y. Shum. Motion texture : A two-level statistical model for character motion synthesis. In *SIGGRAPH*, 2002.

[23] O. Zoeter and T. Heskes. Hierarchical visualization of time-series data using switching linear dynamical systems. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(10):1202–1215, October 2003.