Tractable Maximum Likelihood Estimation for Latent Structure Influence Models with Applications to EEG & ECoG processing

Sajjad Karimi, *Student Member, IEEE,* and Mohammad Bagher Shamsollahi, *Senior Member, IEEE*

Brain signals are nonlinear and nonstationary time series, which provide information about spatiotemporal patterns of electrical activity in the brain. CHMMs are suitable tools for modeling multi-channel time-series dependent on both time and space, but state-space parameters grow exponentially with the number of channels. To cope with this limitation, we consider the influence model as the interaction of hidden Markov chains called Latent Structure Influence Models (LSIMs). LSIMs are capable of detecting nonlinearity and nonstationarity, making them well suited for multi-channel brain signals. We apply LSIMs to capture the spatial and temporal dynamics in multi-channel EEG/ECoG signals. The current manuscript extends the scope of the re-estimation algorithm from HMMs to LSIMs. We prove that the re-estimation algorithm of LSIMs will converge to stationary points corresponding to Kullback-Leibler divergence. We prove convergence by developing a new auxiliary function using the influence model and a mixture of strictly log-concave or elliptically symmetric densities. The theories that support this proof are derived from previous studies by Baum, Liporace, Dempster, and Juang. We then develop a closed-form expression for re-estimation formulas using tractable marginal forward-backward parameters defined in our previous study. Simulated datasets and EEG/ECoG recordings confirm the practical convergence of the derived re-estimation formulas. We also study the use of LSIMs for modeling and classification on simulated and real EEG/ECoG datasets. Based on AIC and BIC, LSIMs perform better than HMMs and CHMMs in modeling embedded Lorenz systems and ECoG recordings. LSIMs are more reliable and better classifiers than HMMs, SVMs and CHMMs in 2-class simulated CHMMs. EEG biometric verification results indicate that the LSIM-based method improves the area under curve (AUC) values by about 6.8% and decreases the standard deviation of AUC values from 5.4% to 3.3% compared to the existing HMM-based method for all conditions on the BED dataset.

***Index Terms*—Coupled Hidden Markov Models, Latent Structure Influence Models, Auxiliary function, Baum-Welch algorithm, Learning problem, EEG.**

I. Introduction

Coupled hidden Markov models (CHMMs) are probabilistic functions that use interacting Markov chains to model complex dynamical systems. CHMMs are used by many applications involving multi-channel time series, such as sign language recognition [1], Audio-Visual Speech Recognition (AVSR) [2], [3], EEG and ECG classification [4]–[6], dynamic Functional Connectivity (dFC) in fMRI [7], disease interactions [8], freeway traffic modeling [9], [10], and financial crisis detection [11]. CHMMs effectively handle non-stationarity and non-linearity that commonly exist in real-world time-series, especially multi-channel brain signals. Multiple brain regions exhibit temporal dependencies, according to brain functional connectivity researches. Therefore a CHMM is a good candidate for analyzing multi-channel brain signals. Study [12] demonstrates a CHMM-based methodology for modeling the trajectory of EEG topography over time. This methodology classifies single trials from visual detection tasks as target and non-target. In a recent study, simulations and real EEG data from epileptic patients were used to test the classification performance [13]. In addition to providing classification results, the model also mapped brain activity back onto the scalp, allowing the EEG signals to be interpreted. Study [14] presents a novel and customized method to detect and localize epileptic seizures in multi-channel scalp EEG recordings. This CHMM framework captures the spatiotemporal propagation for robust seizure detection.

Standard HMMs can model multi-channel interacting time-series, but CHMMs are the better alternatives [15]. They complement the capabilities of standard HMMs by capturing interactions both in space and time for multi-channel time-series [15]. Each channel has its hidden Markov chain and observations (univariate or multivariate) in a C-channel CHMM with N hidden states per channel. Transition probabilities of a channel's hidden states depend on the all previous hidden states, so each channel has a large transition matrix ($N^C \times N$). In general, the number of state-space parameters grows exponentially with the number of channels, and additionally, more sample observations are also needed to estimate them. For example, a 20-channel CHMM with 10 hidden states per channel has 2×10^{22} parameters, and it requires a vast number of observations to learn these parameters that are impossible in practical problems. Therefore the learning problem of CHMMs is much more challenging than HMMs.

There are two common approaches to overcome the exponential growth of state-space parameters in the literature. A simplified factorization of the transition matrix was proposed by Brand in which it was assumed that the probability of a hidden state conditioned on previous states was equal to the product of the marginal conditional probability [15], [16]. While Brand's assumption reduces the number of parameters to $(NC)^2$, it needs N^C normalizing values, as emphasized by [17]. The next approach is the influence model that prevents the exponential growth of state-space parameters [18], [19] , and transition matrices are factorized as follows

$$P(q_t^\xi|q_{t-1}^1,...,q_{t-1}^C) = \sum_{c=1}^{C} \theta^{c,\xi} P(q_t^\xi|q_{t-1}^c),\ \theta^{c,\xi} \geq 0,\ \sum_{c=1}^{C} \theta^{c,\xi} = 1, \tag{1}$$

where, coupling weight $\theta^{c,\xi}$ indicates influence from channel c

M. B. Shamsollahi and S. Karimi are with the Biomedical Signal and Image Processing Laboratory (BiSIPL), School of Electrical Engineering, Sharif University of Technology, Tehran (009821), Iran.
E-mail: mbshams@sharif.edu, sajjadkarimi@ee.sharif.edu

https://github.com/sajjadkarimi91/tractable-mle-lsims will host the codes.

to channel ξ. Coupling weights can be viewed as an adjacency matrix of a graph or network named influence model [19]. The influence model reduces the exponential growth to a quadratic growth $((N^2+1)C^2)$. Several studies used the influence model in higher-order Markov chains modeling, stochastic language modeling, and mixed memory Markov models [18], [20], [21]. Latent Structure Influence Models (LSIMs) are CHMMs with the influence model as the interaction model of Markov chains employed in social computing in several studies [22]–[25].

In contrast to CHMMs, LSIMs can easily be applied to datasets with many channels. Two important problems must be solved for LSIMs to be useful in real-world applications known as inference and learning problems. Exact inference is achieved by transforming an LSIM to an equivalent HMMs with a large cardinality using the Cartesian product of hidden states of all channels. Unfortunately, the exact solution's computational complexity is $\mathcal{O}(TN^{2C})$ [15], [17] that grows exponentially concerning the number of channels. Thus, it can be very demanding and time-consuming, and several approximated inference algorithms were proposed in the literature to overcome this computational demand. Various approximate inference algorithms were proposed to cope with this exponential complexity. The first algorithm uses nonlinear mapping based on Structured Variational Inference (SVI), and marginal forward and backward parameters are obtained by polynomial complexity $\mathcal{O}(T(NC)^2)$ [26]. The next algorithm was developed based on mean-field approximation and variational inference [27], which calculates the one-slice parameter considering the Completely Factorized Variational Inference (CFVI) by computational complexity $\mathcal{O}(T(NC)^3)$. We also proposed a new approximated algorithm to compute marginal forward, backward and one-slice parameters with computational complexity $\mathcal{O}(T(NC)^2)$ [28]. Simulated and real datasets' results confirmed that the proposed inference has less error and superior performance than the previous existing SVI and CFVI.

Learning or estimating LSIM parameters is a more critical and challenging problem than inference. Learning and inference problems are efficiently solved for HMMs to be practical in real-world applications. The well-known forward-backward algorithm solves the inference problem in HMMs. The learning problem involves choosing an optimal set of parameters for some observed multi-channel time series to maximize an appropriate criterion. A well-known training method in HMMs is the Baum-Welch or Expectation-Maximization (EM) algorithm, and it is used to estimate parameters using the maximum-likelihood framework [29]. The EM algorithm finds local maximum likelihood parameters of HMMs based on an auxiliary function, defined upon the Kullback-Leibler divergence [30]. According to Baum's optimization procedure, the optimal parameters are defined as the critical points of the auxiliary functions [31]. Juang extended the EM algorithm of HMMs to accommodate a broad class of mixture of strictly log-concave or elliptically symmetric multivariate distributions [32], [33]. The proof of convergence and closed-form relations of learning in HMMs were presented based on the auxiliary function [32]. A recent study developed a novel Markov chain Monte Carlo (MCMC) algorithm that simultaneously performs inference and parameter estimation in nonhomogeneous Markov chains and puts CHMMs in the context of modeling the spread of infectious diseases [34].

There is no existing learning framework for LSIMs that guarantees the likelihood of model monotonically increases through a re-estimation algorithm and provides proof of convergence. Previous studies developed learning algorithms based on partial derivatives of the likelihood function [17], [22]. The first learning algorithm maximizes a simplified likelihood function with standard constrained optimization [17]. This algorithm uses the chain rule and also takes partial derivatives based on the approximate forward parameter. The re-estimation equations are not explicit as the Baum-Welch algorithm, and channel observations are also considered independent from other channels to simplifying partial derivatives. Thus, there are two sources of error in this framework, and the convergence proof was not discussed. The second learning algorithm is a re-estimation (EM) algorithm developed based on the partial derivatives of a simplified lower band of the log-likelihood function [22], [27]. While this algorithm has a closed-form solution, it does not guarantee a monotonic likelihood increase due to re-estimation procedures. Another study proposed a dynamical influence model for LSIMs with several simplifying assumptions on the structure of transition probabilities and the pattern of coupling weights [25]. A variational EM algorithm is used in the study for the exponential computation of inference, but the theoretical convergence and biases of approximate variational inference are not examined [25].

Thus, the theoretical convergence of the re-estimation algorithm has not been proven exhaustively for LSIMs. Our previous study developed fast and accurate recursive equations to solve approximate inference in LSIMs for a given λ with computational complexity $\mathcal{O}(T(NC)^2)$ [28]. The current study extends the standard EM framework from HMMs to LSIMs and proves monotonic convergence. The current study takes advantage of our previous approximate inference to avoid exponentially high computational demands in maximization re-estimation transforms and proposes a fast and tractable closed-form algorithm for learning LSIM parameters. This work focuses on LSIM learning, and the contributions can be summarized as follows.

- The re-estimation algorithm of HMMs is extended to LSIMs, and convergence is proven theoretically with the influence model and a mixture of strictly log-concave distributions. Convergence is also confirmed practically by applying the proposed algorithm to simulated and real time-series.
- We develop an auxiliary function to prove the convergence for LSIMs theoretically. The technical challenge involves adequately simplifying the observation likelihood based on the influence model. We show that this auxiliary function has a unique global maximum, expressed in a closed-form expression called the re-estimation transformation.
- The re-estimation algorithm is presented as a closed-form expression based on marginal forward-backward parameters similar to the Baum-Welch algorithm. Ap-

proximate inference keeps the complexity at $\mathcal{O}(T(NC)^2)$ instead of $\mathcal{O}(TN^{2C})$, and the proposed re-estimation algorithm works well even for datasets with more than 100 channels.

The rest of this manuscript is structured as follows. In Section II, we describe the notations and constructing an auxiliary function. Then, the re-estimation algorithm's convergence is proved, and the re-estimation algorithm is achieved by maximizing the auxiliary function. In Section III, procedures of data simulation are explained. Several real multi-channel time-series are also introduced, and the validation criteria are described. The proposed re-estimation algorithm is applied to simulated and real multi-channel time-series, and Section IV presents the results of LSIMs comparing to other models. Finally, conclusions are described in Section V, following with expressing some proofs in Appendix B and Appendix C.

II. Proposed Learning Framework

In this section, we define symbols and variables with the same notations as [6], [28]. Some new definitions are also considered for future probabilistic manipulation. Then, a well-established auxiliary function is constructed based on HMMs concepts inspired by the study [32], which is an acceptable source of EM algorithm in HMMs. This function is defined accurately by simplifying $f(o_{1:T}|\lambda)$ based on LSIM parameters. We prove that regarding the influence model, the constructed auxiliary function preserves the structure of the auxiliary function in HMMs. Finally, the uniqueness of the global maximum is showed for the constructed auxiliary function, and then a closed-form solution is also presented to find it. Furthermore, the proposed framework is equivalent to HMM learning framework for an LSIM with one channel.

A. Notations

We assume a LSIM with C channels, and its observations are available for $t = 1, ..., T$. Let us denote $S^c = \{S^c_1, S^c_2, ..., S^c_{M(c)}\}$ to be state space of channel c in the LSIM. Let $q^c_t \in S^c$ and $o^c_t \in \mathbb{R}^{L(c)}$ be state and observation of channel c at time t, respectively and $L(c)$ is observation dimension of channel c. Initial state probabilities of each channel are denoted by $\pi^c_m = P(q^c_1 = S^c_m)$ and $\pi = \{\pi^c_m | m = 1, ..., M(c), c = 1, ..., C\}$. Influence model parameters, including the transition matrices and coupling weights, are denoted by $a^{c,\xi}_{m,n} = P(q^\xi_t = S^\xi_n | q^c_{t-1} = S^c_m)$ and $\theta^{c,\xi}$. The sets of all transition matrices and coupling weights are denoted by A and Θ respectively. Similar to state space, we define channel space as $\Omega_\Theta = \{1, 2, ..., C\}$. The emission probabilities of the observation given its hidden state is written as $b^c_m(o^c_t) = f(o^c_t | q^c_t = S^c_m)$ where o^c_t may be either discrete or continuous. In this study, observations are assumed to be continuous amplitude, and emission probabilities $b^c_m(o^c_t)$ belong to Gaussian Mixture Model (GMM) families as follows

$$\begin{aligned} b^c_m(o^c_t) &= \sum_{k=1}^{D(c)} \omega^c_{m,k} \mathcal{N}(\mu^c_{m,k}, \Sigma^c_{m,k}) \\ &= \sum_{k=1}^{D(c)} \omega^c_{m,k} b^c_{m,k}(o^c_t), \end{aligned} \tag{2}$$

where $D(c)$ is the number of Gaussian in channel c and $\omega^c_m = \{\omega^c_{m,1}, ..., \omega^c_{m,D(c)}\}$, $\mu^c_m = \{\mu^c_{m,1}, ..., \mu^c_{m,D(c)}\}$ and $\Sigma^c_m = \{\Sigma^c_{m,1}, ..., \Sigma^c_{m,D(c)}\}$ are weights, means and covariance matrices of GMM in channel c at state m, respectively. Similar to state space, we define mixture space as $\Omega^c_K = \{1, 2, ..., D(c)\}$. Sets of all mixing weights, means and covariance matrices are also denoted by ω , μ and Σ. Thus, the LSIM is characterized by $\lambda = \{\pi, A, \Theta, \omega, \mu, \Sigma\}$, and Λ is also defined as the total parameters space ($\lambda \in \Lambda$).

Set of observations at time t is denoted by $o_t = \{o^1_t, o^2_t, ..., o^C_t\}$ and set of observations in interval $t_s : t_p$ is denoted by $o_{t_s:t_p} = \{o_{t_s}, o_{t_s+1}, ..., o_{t_p}\}$. A simplifying definition is also considered as $v^c_t(m) \equiv \{q^c_t = S^c_m\}$.

We consider a new variable ϕ^ξ_t in the influence model such that the joint distribution $P(q^\xi_t, \phi^\xi_t | q^1_{t-1}, ..., q^C_{t-1})$ is expressed through

$$P(q^\xi_t, \phi^\xi_t | q^1_{t-1}, ..., q^C_{t-1}) = \theta^{\phi^\xi_t,\xi} P(q^\xi_t | q^{\phi^\xi_t}_{t-1}). \tag{3}$$

The variable ϕ^ξ_t was also defined previously in [21], and this variable indicates the independent partial influence of all channels on channel ξ. The influence model is the marginal distribution of (3) as follows

$$\begin{aligned} P(q^\xi_t | q^1_{t-1}, ..., q^C_{t-1}) &= \sum_{\phi^\xi_t=1}^{C} \theta^{\phi^\xi_t,\xi} P(q^\xi_t | q^{\phi^\xi_t}_{t-1}) \\ \theta^{\phi^\xi_t,\xi} \geq 0, &\quad \sum_{\phi^\xi_t=1}^{C} \theta^{\phi^\xi_t,\xi} = 1. \end{aligned} \tag{4}$$

We further define some sequence spaces in following. Let Ψ^c_Q be the Tth Cartesian product of the hidden states as $\Psi^c_Q = \{(q^c_1, q^c_2, ..., q^c_T) | q^c_t \in S^c, t = 1, ..., T\}$, and the state sequence space of LSIMs is defined as the Cth Cartesian product of Ψ^c_Q ($\Psi_Q = \{\Psi^1_Q, \Psi^2_Q, ..., \Psi^C_Q\}$). Similarly, Ψ^c_K is the Tth Cartesian product of the mixture branches as $\Psi^c_K = \{(\kappa^c_1, \kappa^c_2, ..., \kappa^c_T) | \kappa^c_t \in \Omega^c_K, t = 1, ..., T\}$, and the mixture branch sequence space of LSIMs is defined as the Cth Cartesian product of Ψ^c_K ($\Psi_K = \{\Psi^1_K, \Psi^2_K, ..., \Psi^C_K\}$). Finally, Ψ^c_Φ is the Tth Cartesian product of the channel branches $\Psi^c_\Phi = \{(\phi^c_1, \phi^c_2, ..., \phi^c_T) | \phi^c_t \in \Omega_\Theta, t = 2, ..., T\}$, and the channel branch sequence space of LSIMs is defined as the Cth Cartesian product of Ψ^c_Φ ($\Psi_\Phi = \{\Psi^1_\Phi, \Psi^2_\Phi, ..., \Psi^C_\Phi\}$). A list of acronyms and notations is included in Appendix A to assist the reader in tracking notations in the following sections.

B. Joint Density and Auxiliary Function

The first step of constructing an auxiliary function is to simplify $f(o_{1:T}|\lambda)$ based on LSIM parameters as much as possible. We decompose and simplify the global density function $f(o_{1:T}|\lambda)$ as follows

$$
\begin{aligned}
f(o_{1:T}|\lambda) &= \sum_{\mathcal{Q}\in\Psi_{\mathcal{Q}}} \sum_{K\in\Psi_K} \sum_{\Phi\in\Psi_\Phi} f(o_{1:T},\mathcal{Q},K,\Phi|\lambda) \\
&= \sum_{\mathcal{Q}\in\Psi_{\mathcal{Q}}} \sum_{K\in\Psi_K} \sum_{\Phi\in\Psi_\Phi} \prod_{c=1}^{C} \left(\pi^c_{q^c_1}\omega^c_{q^c_1,\kappa^c_t} b^c_{q^c_1,\kappa^c_1}(o^c_1)\right) \qquad (5) \\
&\times \prod_{t=2}^{T}\prod_{c=1}^{C} \theta^{\phi^c_t,c}\omega^c_{q^c_t,\kappa^c_t} b^c_{q^c_t,\kappa^c_t}(o^c_t) a^{\phi^c_t,c}_{q^{\phi^c_t}_{t-1},q^c_t}.
\end{aligned}
$$

The joint density function $f(o_{1:T},\mathcal{Q},K,\Phi|\lambda)$ is differentiable in λ since the parameters of influence model place in a product form similar to $f(o_{1:T},\mathcal{Q},K|\lambda)$ in HMMs. Following the concept of the Kullback-Leibler divergence in HMMs, we define an auxiliary function $Q(\lambda,\bar{\lambda})$ as a function of two sets of parameters λ and $\bar{\lambda}$ in Λ by the following equation

$$
Q(\lambda,\bar{\lambda}) = \sum_{\mathcal{Q}\in\Psi_{\mathcal{Q}}} \sum_{K\in\Psi_K} \sum_{\Phi\in\Psi_\Phi} f(o_{1:T},\mathcal{Q},K,\Phi|\lambda)\log f(o_{1:T},\mathcal{Q},K,\Phi|\bar{\lambda}). \tag{6}
$$

The next theorem of $Q(\lambda,\bar{\lambda})$ is generalized to LSIMs quickly.

Theorem II.1. *If* $Q(\lambda,\bar{\lambda}) \geq Q(\lambda,\lambda)$ *then* $f(o_{1:T}|\bar{\lambda}) \geq f(o_{1:T}|\lambda)$. *The strict inequality is valid unless* $f(o_{1:T},\mathcal{Q},K,\Phi|\lambda) = f(o_{1:T},\mathcal{Q},K,\Phi|\bar{\lambda})$ *everywhere.*

Proof. The proof follows by Juang and Baum [31], [32]. □

Since (5) and (6) express both the joint density and the auxiliary function, Theorem II.1 is valid. This theorem is essential in the generalized EM algorithm [30]. For a given observed multi-channel time-series $o_{1:T}$, the re-estimation algorithm starts with an initial model λ. Then, a transformation of $\bar{\lambda}$ that increases $Q(\lambda,\bar{\lambda})$ determines the next model in the re-estimation algorithm. However, a better transformation is the maximizer of $Q(\lambda,\bar{\lambda})$ as a function of $\bar{\lambda}$. So, the proposed algorithm re-estimates $\tilde{\lambda}$ from the current model λ as $\tilde{\lambda} = \mathcal{T}(\lambda) \in \{\hat{\lambda} \in \Lambda | Q(\lambda,\hat{\lambda}) = \max_{\bar{\lambda}\in\Lambda} Q(\lambda,\bar{\lambda})\}$. The transformation $\mathcal{T}(\lambda) : \Lambda \rightarrow \Lambda$ is called the re-estimation transformation. Consequently, $\tilde{\lambda}$ plays the same role as λ so that the new re-estimate determines the next model. According to the following theorem, this sequence of models consistently increases $f(o_{1:T}|\lambda)$ unless it reaches a critical point of $f(o_{1:T}|\lambda)$.

Theorem II.2. *Let* $f(o_{1:T},\mathcal{Q},K,\Phi|\lambda)$ *be continuously differentiable in* λ. *If the re-estimation transformation* $\mathcal{T}$ *is defined as a continuous map of* $\Lambda \rightarrow \Lambda$ *such that* $\tilde{\lambda} = \mathcal{T}(\lambda)$ *is a critical point of* $Q(\lambda,\bar{\lambda})$ *as a function of* $\bar{\lambda}$ *, then fixed points of* $\mathcal{T}$ *are critical points of* $f(o_{1:T}|\lambda)$. *Besides, if* $f(o_{1:T}|\tilde{\lambda}) > f(o_{1:T}|\lambda)$, *unless* $\tilde{\lambda} = \lambda$, *all limit points of* $\mathcal{T}^n(\lambda_0) \triangleq \mathcal{T}(\mathcal{T}(...\mathcal{T}(\lambda_0)...))$ *are fixed points of* $\mathcal{T}$ *for any* $\lambda_0 \in \Lambda$.

Proof. The proof follows by Juang and Baum [31], [32]. □

Theorem II.1 and Theorem II.2 guarantee that the model of re-estimation $\tilde{\lambda}$ always increases the likelihood, i.e., $f(o_{1:T}|\tilde{\lambda}) > f(o_{1:T}|\lambda)$, unless $\tilde{\lambda}$ is a fixed point of the transformation. Thus, this transformation will converge to a fixed point, or, in other words, a critical point of the likelihood.

An adequate development of $Q(\lambda,\bar{\lambda})$ may not seems significant until, for example, constructing $Q(\lambda,\bar{\lambda})$ based on $f(o_{1:T},\mathcal{Q}|\lambda)$ instead of $f(o_{1:T},\mathcal{Q},K,\Phi|\lambda)$ also kept Theorem II.1 and Theorem II.2 going. Two essential consideration of constructing a new $Q(\lambda,\bar{\lambda})$ should be noticed: the uniqueness of its global maximum and a fast closed-form solution of this global maximum. Next, we indicate that these points exist in the proposed $Q(\lambda,\bar{\lambda})$.

C. Maximization and Re-estimation Algorithm

The logarithm of global joint density has a separability property as follows

$$
\begin{aligned}
\log f(o_{1:T},\mathcal{Q},K,\Phi|\bar{\lambda}) &= \sum_{c=1}^{C}\log\bar{\pi}^c_{q^c_1} + \sum_{c=1}^{C}\sum_{t=2}^{T}\log\bar{\theta}^{\phi^c_t,c} \\
+\sum_{c=1}^{C}\sum_{t=2}^{T}\log\bar{a}^{\phi^c_t,c}_{q^{\phi^c_t}_{t-1},q^c_t} &+ \sum_{c=1}^{C}\sum_{t=1}^{T}\log\bar{\omega}^c_{q^c_t,\kappa^c_t} + \sum_{c=1}^{C}\sum_{t=1}^{T}\log\bar{b}^c_{q^c_t,\kappa^c_t}(o^c_t).
\end{aligned} \tag{7}
$$

Assuming there are I separable parameter sets such that $\log f(o_{1:T},\mathcal{Q},K,\Phi|\lambda) = \sum_{i=1}^{I}\log f^{(i)}(o_{1:T},\mathcal{Q},K,\Phi|\lambda_i)$. Note that $\lambda = \{\lambda_1,...,\lambda_I\}$ and I is the number of parameter sets after separation. Define $Q(\lambda,\bar{\lambda}_i)$ like this

$$
\begin{aligned}
&Q(\lambda,\bar{\lambda}_i) = \\
&\sum_{\mathcal{Q}\in\Psi_{\mathcal{Q}}} \sum_{K\in\Psi_K} \sum_{\Phi\in\Psi_\Phi} f(o_{1:T},\mathcal{Q},K,\Phi|\lambda)\log f^{(i)}(o_{1:T},\mathcal{Q},K,\Phi|\bar{\lambda}_i).
\end{aligned} \tag{8}
$$

Assuming λ is fixed and $Q(\lambda,\bar{\lambda}_i)$ as a function of $\bar{\lambda}_i$ has a unique global maximum $\tilde{\lambda}_i$, that is a critical point of $Q(\lambda,\bar{\lambda}_i)$. The partial transformation is defined as $\mathcal{T}^{(i)} : \lambda \rightarrow \breve{\lambda}_i = \{\lambda_1,...,\tilde{\lambda}_i,...,\lambda_I\}$, and the re-estimation transformation $\mathcal{T}$ is thus defined as $\mathcal{T} : \lambda \rightarrow \tilde{\lambda} = \{\tilde{\lambda}_1,...,\tilde{\lambda}_i,...,\tilde{\lambda}_I\}$. Based on the following theorem, maximizing the likelihood through re-estimation can be performed on individual parameter sets.

Theorem II.3. *According to the assumptions outlined above,* $f(o_{1:T}|\mathcal{T}^{(i)}(\lambda)) \geq f(o_{1:T}|\lambda)$ *for all* $\lambda \in \Lambda$, *and every parameter set* i. *The inequality becomes equality if and only if* λ_i *is a critical point of* $f(o_{1:T}|\lambda)$ *with regard to* λ_i, *or equivalently,* $\breve{\lambda}_i$ *is a fixed point of* $\mathcal{T}^{(i)}$. *Besides,* $f(o_{1:T}|\mathcal{T}(\lambda)) \geq f(o_{1:T}|\lambda)$ *with equality if and only if* λ *is a critical point of* $f(o_{1:T}|\lambda)$ *or equivalently, a fixed point of* $\mathcal{T}$.

Proof. The proof follows by Juang and Baum [31], [32]. □

Now, we decompose $Q(\lambda,\bar{\lambda})$ as the sum of separated parts by substituting (7) in (6). It is then proved that each separated part has a unique global maximum (including influence model parameters).

Let $\boldsymbol{\Upsilon}^c_{n,\kappa}$ is the parameter set that defines the density $b^c_{n,\kappa}(o^c_t)$. This study assumes multivariate Gaussian densities and $\boldsymbol{\Upsilon}^c_{n,\kappa} = (\mu^c_{n,\kappa},\Sigma^c_{n,\kappa})$. Inserting (7) into the auxiliary function in (6) gives

$$
\begin{aligned}
Q(\lambda,\bar{\lambda}) = \sum_{c=1}^{C} Q_{\bar{\boldsymbol{\pi}}^c}(\lambda,\bar{\boldsymbol{\pi}}^c) + \sum_{c=1}^{C} Q_{\bar{\boldsymbol{\theta}}^c}(\lambda,\bar{\boldsymbol{\theta}}^c) \\
+ \sum_{c=1}^{C}\sum_{n_c=1}^{M(c)} Q_{\bar{\boldsymbol{\omega}}^c_{n_c}}(\lambda,\bar{\boldsymbol{\omega}}^c_{n_c}) + \sum_{c=1}^{C}\sum_{n_c=1}^{M(c)}\sum_{\kappa_c=1}^{D(c)} Q_{\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c}}(\lambda,\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c}) \\
+ \sum_{c=1}^{C}\sum_{c'=1}^{C}\sum_{n_{c'}=1}^{M(c')} Q_{\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}}(\lambda,\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}),
\end{aligned}
\tag{9}
$$

where $Q(\lambda,\bar{\lambda}_i)$ is described in detailed as follows

$$
\begin{aligned}
Q_{\bar{\boldsymbol{\pi}}^c}(\lambda,\bar{\boldsymbol{\pi}}^c) = \sum_{n_c=1}^{M(c)} f(o_{1:T}, v_1^c(n_c)|\lambda)\log\bar{\pi}^c_{n_c} \\
Q_{\bar{\boldsymbol{\theta}}^c}(\lambda,\bar{\boldsymbol{\theta}}^c) = \sum_{c'=1}^{C}\log\bar{\theta}^{c',c}\sum_{t=2}^{T} f(o_{1:T},\phi_t^c = c'|\lambda) \\
Q_{\bar{\boldsymbol{\omega}}^c_{n_c}}(\lambda,\bar{\boldsymbol{\omega}}^c_{n_c}) = \sum_{k_c=1}^{D(c)}\log\bar{\omega}^c_{n_c,k_c}\sum_{t=1}^{T} f(o_{1:T}, v_t^c(n_c), \kappa_t^c = k_c|\lambda) \\
Q_{\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c}}(\lambda,\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c}) = \sum_{t=1}^{T} f(o_{1:T}, v_t^c(n_c), \kappa_t^c = k_c|\lambda)\log\bar{b}^c_{q_t^c,k_c}(o_t^c) \\
Q_{\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}}(\lambda,\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}) = \\
\sum_{n_c=1}^{M(c)}\log\bar{a}^{c',c}_{n_{c'},n_c}\times\sum_{t=2}^{T} f(o_{1:T},\phi_t^c = c', v_{t-1}^{\phi_t^c}(n_{c'}), v_t^c(n_c)|\lambda).
\end{aligned}
\tag{10}
$$

Under the Theorem II.3, if each individual auxiliary function has a unique maximum global, then the parameter sets can be re-estimated independently by maximizing the individual auxiliary functions independently. Fortunately, maximization of $Q_{\bar{\boldsymbol{\pi}}^c}(\lambda,\bar{\boldsymbol{\pi}}^c)$, $Q_{\bar{\boldsymbol{\theta}}^c}(\lambda,\bar{\boldsymbol{\theta}}^c)$, $Q_{\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}}(\lambda,\bar{\boldsymbol{a}}^{c',c}_{n_{c'}})$ and $Q_{\bar{\boldsymbol{\omega}}^c_{n_c}}(\lambda,\bar{\boldsymbol{\omega}}^c_{n_c})$ subject to the following constraints is well-known (for all appropriate c, c', and n_c)

$$
\begin{aligned}
\sum_{n_c=1}^{M(c)}\bar{\pi}^c_{n_c} = 1, \quad \bar{\pi}^c_{n_c} \ge 0 \\
\sum_{c'=1}^{C}\bar{\theta}^{c',c} = 1, \quad \bar{\theta}^{c',c} \ge 0 \\
\sum_{n_c=1}^{M(c)}\bar{a}^{c',c}_{n_{c'},n_c} = 1, \ \ \bar{a}^{c',c}_{n_{c'},n_c} \ge 0 \\
\sum_{k_c=1}^{D(c)}\bar{\omega}^c_{n_c,k_c} = 1, \ \ \bar{\omega}^c_{n_c,k_c} \ge 0.
\end{aligned}
\tag{11}
$$

Each auxiliary function has a well-known form $\sum_{j=1}^{N} w_j\log y_j$ coupled with the constraints $\sum_{j=1}^{N} y_j = 1$, $y_j \ge 0$ and $w_j \ge 0$, which leads to a unique global maximum as follows [32]

$$
y_j = \frac{w_j}{\sum_{n=1}^{N} w_n}
\tag{12}
$$

Hence, $Q_{\bar{\boldsymbol{\pi}}^c}(\lambda,\bar{\boldsymbol{\pi}}^c)$, $Q_{\bar{\boldsymbol{\theta}}^c}(\lambda,\bar{\boldsymbol{\theta}}^c)$, $Q_{\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}}(\lambda,\bar{\boldsymbol{a}}^{c',c}_{n_{c'}})$, and $Q_{\bar{\boldsymbol{\omega}}^c_{n_c}}(\lambda,\bar{\boldsymbol{\omega}}^c_{n_c})$ have a unique global maximum. In this study, $\bar{b}^c_{q_t^c,\kappa_t^c}(o_t^c)$ is considered as a Gaussian distribution, which belongs to elliptically symmetric distributions. Thus, $Q_{\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c}}(\lambda,\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c})$ also has a unique global maximum since $\bar{b}^c_{q_t^c,\kappa_t^c}(o_t^c)$ is elliptically symmetric [33]. Therefore, all individual auxiliary functions have a unique global maximum, and parameters of LSIMs can be estimated by iterative maximization of individual auxiliary functions according to the mentioned theorems.

The re-estimation transformation is derived by applying (12) to individual auxiliary functions $Q_{\bar{\boldsymbol{\pi}}^c}(\lambda,\bar{\boldsymbol{\pi}}^c)$, $Q_{\bar{\boldsymbol{\theta}}^c}(\lambda,\bar{\boldsymbol{\theta}}^c)$, $Q_{\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}}(\lambda,\bar{\boldsymbol{a}}^{c',c}_{n_{c'}})$, and $Q_{\bar{\boldsymbol{\omega}}^c_{n_c}}(\lambda,\bar{\boldsymbol{\omega}}^c_{n_c})$. Maximization of $Q_{\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c}}(\lambda,\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c})$ is also well-known and straightforward [33].

We have a brief review of inference in LSIMs that including forward, backward, one-slice parameters. Then, we use these parameters to accomplish the re-estimation transformation by maximization of individual auxiliary functions.

1) Inference in LSIMs and Auxiliary Parameters

Inference in an LSIM is to compute the conditional probabilities of hidden states at a time t given some duration of observations. The forward, backward, and one-slice parameters are the most critical parameters in the inference that have an essential role in re-estimation transformation. The marginal forward, one-slice, and backward parameters are respectively defined as follows [28]

$$
\begin{aligned}
\alpha^c_{t|t-1}(m) = P(v_t^c(m)|o_{1:t-1}) \\
\alpha^c_{t|T}(m) = P(v_t^c(m)|o_{1:T}) \\
\beta_t^c(m) = \frac{\alpha^c_{t|T}(m)}{\alpha^c_{t|t-1}(m)} = \frac{b^c_m(o_t^c)}{f(o_t^c|o_{1:t-1})}\times\frac{f(o_{t+1:T}|v_t^c(m),o_{1:t})}{f(o_{t+1:T}|o_{1:t})}.
\end{aligned}
\tag{13}
$$

To complete the maximization process, two new auxiliary parameters $\Gamma_t^{c',c}(n_c',n_c)$ and $\gamma_t^c(n_c,k_c)$, must also be defined that expressed according to the previous forward, backward and one-slice parameters. Two-slice parameter $\Gamma_t^{c',c}(n_c',n_c)$ is the joint distribution of two consecutive states (from different or same channels) plus ϕ_t^c given all observation. This parameter is defined and simplified as follows (see Appendix B)

$$
\begin{aligned}
\Gamma_t^{c',c}(n_{c'},n_c) = P(\phi_t^c = c', v_{t-1}^{\phi_t^c}(n_{c'}), v_t^c(n_c)|o_{1:T},\lambda) \\
= \alpha^{c'}_{t-1|t-1}(n_{c'})\theta^{c',c}a^{c',c}_{n_{c'},n_c}\beta_t^c(n_c).
\end{aligned}
\tag{14}
$$

The next parameter is $\gamma_t^c(n_c,k_c)$, which is defined as the joint distribution of the hidden state and its mixture branch conditioned on all observation. This parameter is expressed as follows (see Appendix B)

$$
\begin{aligned}
\gamma_t^c(n_c,k_c) = P(v_t^c(n_c),\kappa_t^c = k_c|o_{1:T},\lambda) \\
= \frac{\omega^c_{n_c,k_c}b^c_{n_c,k_c}(o_t^c)}{\sum_{\kappa=1}^{D(c)}\omega^c_{n,\kappa}b^c_{n,\kappa}(o_t^c)}\alpha^c_{t|T}(n_c).
\end{aligned}
\tag{15}
$$

2) Maximization of Individual Auxiliary Functions

Here, introduced parameters are used to extract partial re-estimation transformations achieved by maximizing individual auxiliary functions. Maximization procedures of individual auxiliary functions are described in detail in Appendix C. Initial probabilities are re-estimated by the maximization of $Q_{\bar{\boldsymbol{\pi}}^c}(\lambda,\bar{\boldsymbol{\pi}}^c)$ as follows

$$\tilde{\pi}^{c}_{n_c} = \alpha^{c}_{1|T}(n_c). \tag{16}$$

Coupling weights and transition matrices of the influence model are also re-estimated by maximizing $Q_{\bar{\boldsymbol{\theta}}^c}(\lambda, \bar{\boldsymbol{\theta}}^{c})$ and $Q_{\bar{\boldsymbol{a}}^{c',c}_{n_{c'}}}(\lambda, \bar{\boldsymbol{a}}^{c',c}_{n_{c'}})$ through the following equations

$$\tilde{\theta}^{c',c} = \frac{\sum_{n_c=1}^{M(c)} \sum_{n_{c'}=1}^{M(c')} \sum_{t=2}^{T} \Gamma_t^{c',c}(n_{c'}, n_c)}{\sum_{s=1}^{C} \sum_{n_c=1}^{M(c)} \sum_{n_s=1}^{M(s)} \sum_{t=2}^{T} \Gamma_t^{s,c}(n_s, n_c)} \tag{17}$$

$$\tilde{a}^{c',c}_{n_{c'},n_c} = \frac{\sum_{t=2}^{T} \Gamma_t^{c',c}(n_{c'}, n_c)}{\sum_{n=1}^{M(c)} \sum_{t=2}^{T} \Gamma_t^{c',c}(n_{c'}, n)}. \tag{18}$$

Maximization of $Q_{\bar{\boldsymbol{\omega}}^c_{n_c}}(\lambda, \bar{\boldsymbol{\omega}}^{c}_{n_c})$ also gives mixing weights of GMMs according to

$$\tilde{\omega}^{c}_{n_c,k_c} = \frac{\sum_{t=1}^{T} \gamma_t^c(n_c, k_c)}{\sum_{\kappa=1}^{D(c)} \sum_{t=1}^{T} \gamma_t^c(n_c, \kappa)} \tag{19}$$

Lastly, the mean vector and covariance matrix of GMMs are re-estimated by the maximization of $Q_{\bar{\boldsymbol{\Upsilon}}^c_{n_c,k_c}}(\lambda, \bar{\boldsymbol{\Upsilon}}^{c}_{n_c,k_c})$ as follows

$$\tilde{\mu}^{c}_{n_c,k_c} = \frac{\sum_{t=1}^{T} \gamma_t^c(n_c, k_c) o_t^c}{\sum_{t=1}^{T} \gamma_t^c(n_c, k_c)}$$
$$\tilde{\Sigma}^{c}_{n_c,k_c} = \frac{\sum_{t=1}^{T} \gamma_t^c(n_c, k_c)(o_t^c - \mu^c_{n_c,k_c})(o_t^c - \mu^c_{n_c,k_c})^\top}{\sum_{t=1}^{T} \gamma_t^c(n_c, k_c)}. \tag{20}$$

Above closed-form solutions (partial re-estimation transformations) are impractical for a dataset with many channels due to the exponentially computational complexity of exact marginal parameters. Our previous study proposed a fast approximate inference that computes marginal parameters fast and recursively with the computational complexity $\mathcal{O}(T(NC)^2)$ instead of $\mathcal{O}(TN^{2C})$ for an LSIM with C channels of N states apiece observing T data points [28]. This approximate algorithm is fast and acceptable for many practical applications, while exact inference can be demanding and time-consuming. Hellinger distances are small enough, indicating that the proposed approximate inference is sufficiently close to the exact inference when considering various channels, hidden states, and other parameters [28]. Further, the proposed inference algorithm has superior performance than existing approximate inference algorithms. Therefore, our proposed forward, backward, and one-slice parameters are used in partial re-estimation transformation in the remainder of this manuscript.

III. Simulated and Real Datasets

This section describes various simulated and EEG/ECoG datasets to evaluate the proposed learning algorithm. Then, we consider the application of multi-channel time-series modeling using LSIMs and HMMs. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used to compare HMMs and LSIMs in the context of modeling on both simulated and real datasets. Since one of the most important applications of LSIMs as a generative model is in time-series classification [35], [36], we also evaluate the performance of LSIMs against CHMMs, HMMs and Support Vector Machines (SVMs).

A. Simulated Datasets

1) Generic CHMMs

The first simulation includes a generic CHMM that generates simulated multi-channel time-series. Parameters of a generic CHMM denote by $\lambda = \{\pi, A_g, \omega, \mu, \Sigma\}$. Notice that $A_g = \{A^1_g, ..., A^C_g\}$, where $A^c_g \in \mathbb{R}^{N^C \times N}$ in CHMMs refers to transition probabilities different from those of the influence model in LSIMs. Consider a CHMM with C channels; each channel takes a random state number between 2 to 6. Initial state probabilities $\pi^c = \{\pi^c_1, ..., \pi^c_{M(c)}\}$ is initialized by uniform distribution $\mathcal{U}(0, 1)$, and then normalized dividing them to the sum of them. In the same way, each row of A^c_g is also initialized and normalized. The observation dimension of each channel also initialized randomly between 1 to 5. Emission probabilities belong to GMM generally, but it is simpler to assume just one Gaussian component. $\mathcal{N}(m, 1)$ initializes each element of the mean vector $\mu^c_{m,k}$. For simplicity, covariance matrices of emission probabilities are assumed to be diagonal, and $\mathcal{U}(1, 3)$ initializes their diagonal elements.

2) Embedded Lorenz Systems

Lorenz system is an interesting nonlinear dynamical equation with a set of ordinary differential equations known as Lorenz equations

$$\begin{cases} \frac{dx}{dt} = \sigma(y - x) \\ \frac{dy}{dt} = x(\rho - z) - y \\ \frac{dz}{dt} = xy - \beta z. \end{cases} \tag{21}$$

Hundreds of research articles and at least one book-length studied the Lorenz equations [37]. Two different sets ($\{\sigma = 10, \rho = \frac{8}{3}, \beta = 28\}$ and $\{\sigma = 10, \rho = \frac{8}{3}, \beta = 56\}$) were selected to generate time-series of the Lorenz system. The system exhibits chaotic behavior for these sets. Each set of parameters generated a three-channel time-series, and these 3-channel time-series are embedded to create a new six-channel time-series. The proposed algorithm is applied to a six-channel time-series that has more complexity than a three-channel case.

B. Brain Signal Datasets

Besides simulated datasets, we consider EEG and ECoG datasets to assess performances better, as described in the following.

1) EPFL EEG dataset

This dataset comprises EEG recordings of five disabled and four healthy subjects [38]. EEG was recorded from 32 electrodes placed at the standard positions of the 10-20 international system. The initial sampling rate was 2048 Hz that was resampled to 128 Hz. Subjects were facing a laptop screen on which six images were displayed. The images showed a television, a telephone, a lamp, a door, a window, and a radio. The dataset and preprocessing used in the present work are made available for download on the EPFL BCI group. The recording of subject five was excluded from the dataset due to the presence of artifacts. Study [38] had described the detail of protocols and subjects.

2) *Macaques ECoG dataset*

This dataset contains ECoG recordings of monkeys in a tracking food task [39]. ECoG (64-channel) and hand motion data were recorded simultaneously during the tracking food task with sampling rates 1 kHz and 120 Hz, respectively. This study focused on monkey B recording to have more summarized results due to space limitation.

3) *Biometric EEG Dataset (BED)*

The purpose of this dataset is to study the performance of methods for the task of biometric person verification and identification [40]. BED is a dataset designed for EEG-based biometrics, using a low-cost consumer-grade EPOC+ measuring 14-channels of EEG using contact sensors located at locations that closely align with the AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4 locations of the international 10-20 system. BED contains EEG recordings acquired from 21 healthy individuals with 12 different types of stimuli. EEG signals are collected throughout three sessions spaced one week apart to study template aging. The experimental protocol consists of affective stimuli (AS), mathematical computations (MC), resting eyes closed (RC), resting eyes open (RO) and visual evoked potentials (VEP) at 3, 5, 7, and 10 Hz with a standard checkerboard pattern with reversed pattern (VCx, x = 3; 5; 7; 10), and flashing VEP with flashing black color at 3, 5, 7, and 10 Hz (VFx, x = 3; 5; 7; 10).

IV. Results and Discussion

This section presents the results of applying LSIMs to simulated and real datasets in well-categorized parts. We first check convergence, then examine modeling and classification applications, and next analyze approximate inference biases. We select a subset of datasets for each part to create more summarized and comprehensive results.

A. *Convergence Testing*

The proposed learning algorithm is applied to various multi-channel time-series to examine its convergence and monotonically increasing.

1) *Simulated data*

Convergence is tested first using simulated multi-channel time series. Fig. 1 displays log-likelihood curves for 3 and 5 channels simulated observations of CHMMs. All channels are in similar states, and the number of states increases from 2 to 7 in the re-estimation algorithm. In all cases, the log-likelihood curves increase monotonically, as shown in the figure. Results show that the final log-likelihood increases as the number of states in LSIMs increases. Furthermore, simple models tend to have faster convergence at the same number of iterations.

2) *EPFL EEG dataset*

We further apply the proposed algorithm to the EPFL EEG dataset. Two different configurations are analyzed, including 16 and 32 channels with the same channel selection as [38]. In all subjects, the state number is four with two Gaussian components. Fig. 2 shows log-likelihood curves that are consistent with previous simulation results.

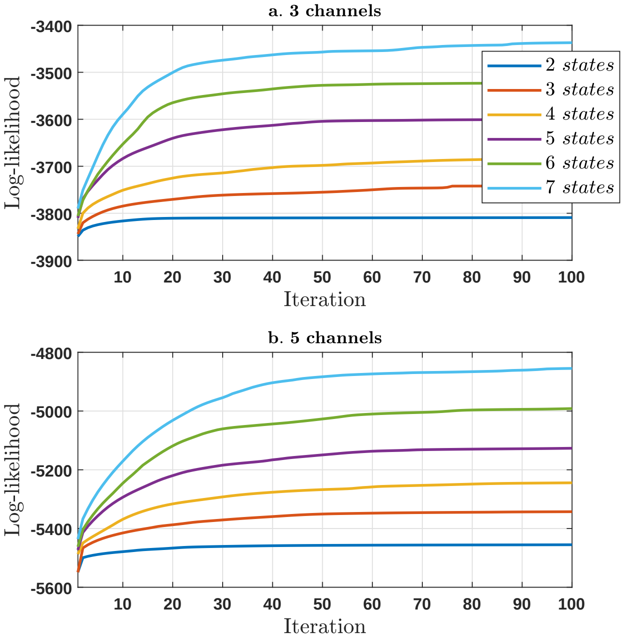


Fig. 1. Convergence of proposed algorithm in the estimation of LSIMs parameters from simulated multi-channel time-series for 3 and 5 channels CHMMs

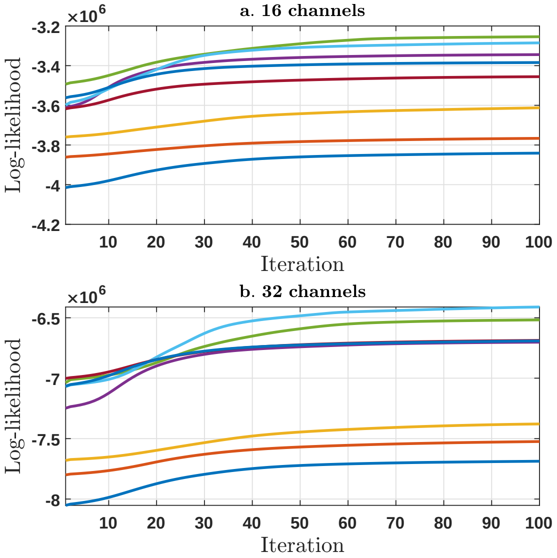


Fig. 2. Convergence of proposed EM-algorithm in the estimation of LSIM parameters considering the various number of real EEG channels and subjects

3) *Macaque ECoG dataset*

In the end, the Macaque ECoG dataset with 64 channels confirms the convergence of the proposed algorithm in a high channel number scenario. Fig. 3 shows the log-likelihood curves for different state numbers in this dataset. Log-likelihood curves increase monotonically in all cases, indicating that the proposed algorithm has a stable monotonic convergence.

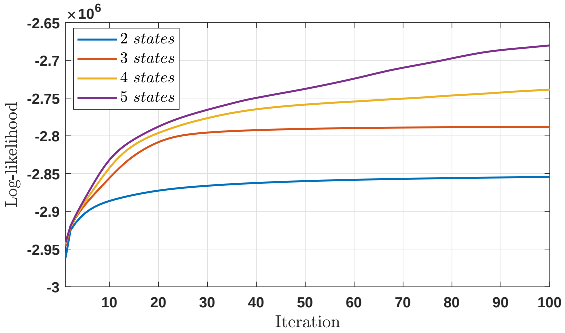


Fig. 3. Convergence of proposed algorithm in the estimation of LSIM parameters considering a 64-channel ECoG

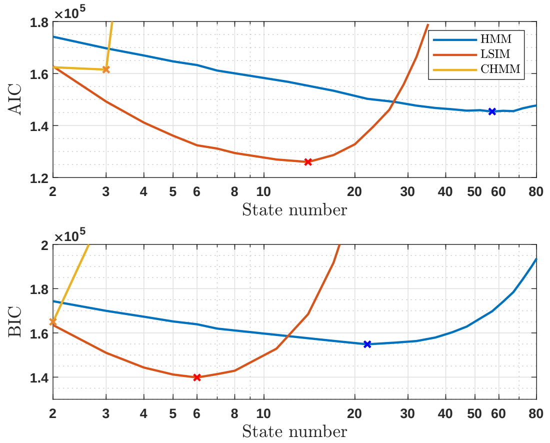


Fig. 4. Analysis of AIC and BIC for HMM, CHMM and LSIM considering embedded Lorenz systems

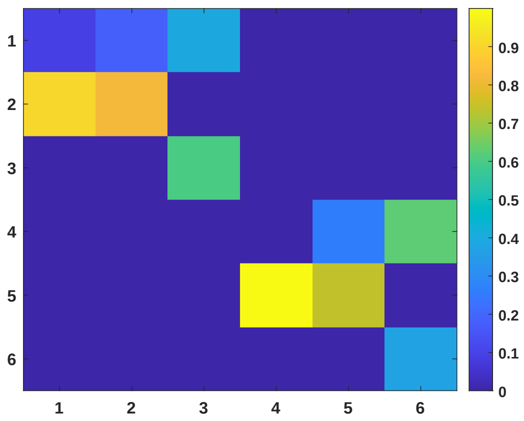


Fig. 5. Coupling weights of the selected model based on BIC for embedded Lorenz system

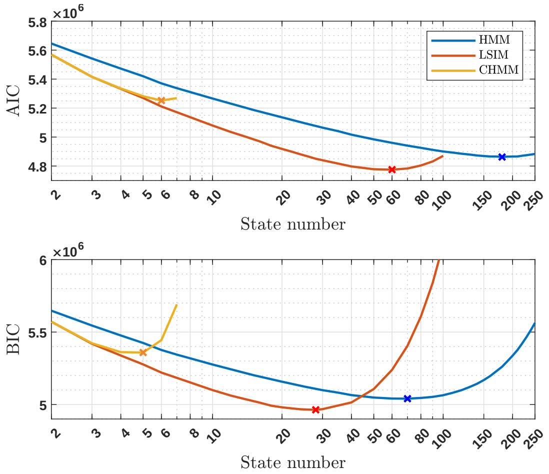


Fig. 6. Analysis of AIC and BIC for HMM, CHMM and LSIM considering multichannel ECoG

B. *Statistical Modeling of Multi-channel Signals*

In addition to checking convergence, simulated embedded Lorenz systems and the Macaque ECoG dataset are used to compare the modeling capability of LSIMs with HMMs.

1) *Embedded Lorenz Systems*

The first scenario involves applying HMMs, CHMMs and LSIMs to a six-channel time series from embedded Lorenz systems. A simple grid-search also finds optimal state numbers (2-35 for LSIMs, 2-80 for HMMs and 2-4 for CHMMs with one Gaussian component) based on AIC and BIC. LSIMs and CHMMs can have different states per channel, and the exact grid-search grows exponentially with the number of channels. To avoid this exponential grid-search, we assume that all channels have the same number of states. The AIC and BIC curves for HMMs, CHMMs and LSIMs are shown in Fig. 4. Based on AIC and BIC, LSIMs are better than HMMs and CHMMs. In addition, Fig. 5 shows the coupling weights of the selected LSIM based on the BIC, and as expected, the coupling weights are zero between two independent Lorenz systems.

2) *Macaque ECoG dataset*

Macaque ECoG dataset contains 64 recording electrodes, which can be modeled by a 64-channel LSIM (one electrode per channel), a very complex model with lots of parameters. This complexity can be reduced by considering a four-channel LSIM consisting of 16 electrodes per channel such that each channel contains 16 neighbor electrodes on the multi-electrode array (see [39] for more details). The cross-correlation matrix also supports this reconfiguration (see Fig. 7.a). The covariance matrices in GMMs are assumed to be diagonal matrices to avoid singularities [41].

Similar to the previous dataset, a simple grid-search is used to obtain the optimal state number (2-100 for LSIMs, 2-250 for HMMs and 2-7 for CHMMs with one Gaussian component). Fig. 6 indicates that LSIMs perform better in terms of AIC and BIC than HMMs and CHMMs. Moreover, Fig. 7.b shows the coupling weights of the selected model according to the BIC, consistent with the cross-correlation matrix.

3) *Computational Efficiency*

One of the advantages of the proposed LSIM framework is its computational efficiency in large channel systems where CHMMs cannot be used due to their exponential computation requirements. Model learning computation times are reported in TABLE I for LSIMs compared to CHMMs when using the ECoG and embedded Lorenz datasets. This table presents results for a four-channel ECoG dataset with hidden states (M)

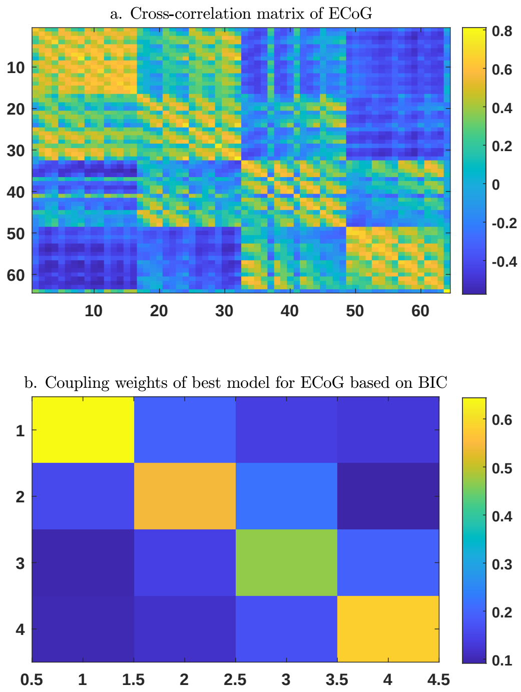


Fig. 7. Coupling weights of the selected model based on BIC for Macaque ECoG dataset

TABLE I
LSIM AND CHMM LEARNING COMPUTATION TIMES (SECONDS) FOR HIDDEN STATES (M) PER CHANNEL VARYING FROM 2 TO 4

	ECoG ($C = 4$)			Lorenz ($C = 6$)		
	M=2	M=3	M=4	M=2	M=3	M=4
LSIM	51	60	84	19	20	22
CHMM	37	146	553	15	746	25394

varying from 2 to 4. When M is increased from 2 to 4, the LSIMs computational times increase from 51 to 84 seconds, while the CHMMs computational times increase quickly from 37 to 553 seconds. A six-channel embedded Lorenz learning computation takes 19 to 22 seconds, while the CHMMs learning computation takes 15 to 25394 seconds. According to the table, CHMMs computational times increase exponentially with the number of channels and states, making it impractical for real-world datasets. In contrast, the computational time of LSIMs remains appropriate for real-world datasets with many channels.

C. Classification & Biometric Verification

LSIMs can also classify multi-channel time-series data, as HMMs and CHMMs can do. If there are different generative models behind different time series classes, HMMs, CHMMs and LSIMs may be more appropriate than other classifiers [35]. During the training phase of a two-class classification problem, two Markov models are learned using train observations from each class. Hence, there are two models with λ_1 and λ_2 parameters, and a given test time series must be assigned to one of these models. For this assignment, the conditional observation likelihood of the test time series is computed for model parameters λ_1 and λ_2, and ll_{λ_1} and ll_{λ_2} indicate their log-likelihoods. If $ll_{\lambda_1} - ll_{\lambda_2} > 0$, then the test time series is assigned to model λ_1 and vice versa. This part compares the accuracy of LSIMs against HMMs, CHMMs, and SVMs using simulated CHMMs and real EEG datasets.

TABLE II
CLASSIFICATION ACCURACY FOR SVM, HMM, CHMM AND LSIM IN 3-CHANNEL SIMULATED CHMMS

	$I = 10$		$I = 15$		$I = 20$	
	T=5	T=10	T=5	T=10	T=5	T=10
SVM	71.1	74.7	73.9	78.8	75.6	80.8
HMM	76.1	85.8	79.8	89.7	82.3	91.5
LSIM	82.6	**93.9**	**87.7**	**96.3**	**90**	**97.2**
CHMM	**83.7**	93.7	87.1	95.7	88.8	96.5

1) Simulated CHMMs

Under the classification scenario, two CHMMs parameters are initialized randomly with the same structure (same channel numbers, channel dimensions, and state number per channel) to generate train and test observations for a two-class problem.

HMMs, CHMMs, and LSIMs have two hyper-parameters, including hidden state numbers and the number of Gaussian components. A proper model selection criterion (AIC, BIC, ...) must be used to determine the optimal hyper-parameters based on the training observations without needing a validation set. The optimal hyper-parameters are determined by minimizing AIC using a grid search on hidden state and Gaussian components. The optimal hyper-parameters are then applied to the test time series to classify them.

We analyze the effect of channel number (C), the number of training set samples (I), and sequence duration (T) on the classification accuracy. There are 10 test time-series with the same T as the training set for each CHMM, and CHMM parameters are reinitialized 1000 times for each particular condition (including C, I, and T). Thus, the classification accuracy per condition is expressed in terms of 20000 test time-series.

TABLE II and TABLE III present classification accuracy for various C, I, and T. The results show that increasing C, I, or T positively affects classification accuracy independent of the LSIMs, CHMMs, HMMs, or SVMs, which are equivalent to increasing the given information.

The tables show that LSIMs are superior to HMMs, CHMMs and SVMs, and CHMMs perform better than HMMs and SVMs. TABLE II indicates LSIMs are better than CHMMs by about 0.4% on average for the 3-channel simulation scenario, and TABLE III shows LSIMs beat CHMMs by about 3% for the 5-channel simulation scenario. This improvement shows that LSIMs become more appropriate and stronger than CHMMs by increasing the number of channels in interesting datasets, even for simulated data from generic CHMMs.

2) BED biometric verification

This part aims to improve the verification performance of the BED dataset by using LSIMs instead of HMMs for EEG-based person verification. The verification task is to determine whether a user is who they claim to be. Verification compares the query with the template of the requested identity, and

TABLE III
CLASSIFICATION ACCURACY FOR SVM, HMM, CHMM AND LSIM IN 5-CHANNEL SIMULATED CHMMS

	$I = 10$		$I = 15$		$I = 20$	
	T=5	T=10	T=5	T=10	T=5	T=10
SVM	76.7	82.2	79.1	84.8	81.7	86.7
HMM	83.3	92.9	85.7	94.8	88.7	95.9
LSIM	**88.3**	**98.3**	**92.7**	**99**	**95.2**	**99.3**
CHMM	84.3	95.8	88.5	97.3	91.5	97.7

users are accepted or rejected based on whether the result of the comparison exceeds or falls below a certain threshold. In contrast, identification refers to deciding who the user is from a pool of possible profiles. This context compares all available profiles and assigns the query to the profile that provides the best match.

In [40], an HMM-based verification method is developed, and verification performance is evaluated on the BED dataset for different types of features and stimuli. A similar EEG-based verification is also conducted using LSIMs as in [40] to avoid ambiguity or questions on the processing procedures. The current study follows the same data preprocessing and epoching, feature extraction, and the decision rule for accepting or rejecting an epoch proposed in [40].

In summary, the recordings of channel c data ($c = 1, 2, .., C$) are segmented into P consecutive 5-second epochs with 50% overlap ($e^{(c,p)}, p = 1, 2, .., P$). Next, epoch p is split into H overlapping frames of 1 second and 50% overlapping, and represented as a sequence of observations $o^{(c,p)}$, so that $o^{(c,p)} = [f_1^{(c,p)}, f_2^{(c,p)}, ..., f_H^{(c,p)}]$, where $f_h^{(c,p)}$ denotes the h-th frame of epoch $e^{(c,p)}$. Every frame is then used to extract the feature vector $\hat{f}_h^{(c,p)}$ that contains F features. Mel frequency cepstral coefficients (MFCC), autoregression reflection coefficients (ARRC), and spectral features (SPEC) are extracted for each frame with F_{MFCC}=12, F_{ARRC}=12 and F_{SPEC}=14. The observation sequence in the feature space is denoted as $\hat{o}^{(c,p)} = [\hat{f}_1^{(c,p)}, \hat{f}_2^{(c,p)}, ..., \hat{f}_H^{(c,p)}]$. The dataset of features is also part of the BED dataset, and we directly downloaded the preprocessed features for verification evaluation.

Existing works build an HMM λ^c_{HMM} with 4 hidden states using $\hat{o}^{(c,p)}$, then compute a posteriori log likelihood $l^{(c,p)} = P(\hat{o}^{(c,p)}|\lambda^c_{HMM})$ for EEG channel c with respect to the maximum probability path through the Viterbi algorithm [40], [42]. The decision rule for accepting or rejecting epoch p based on C models of any given subject is according to [40], [42]

$$z_p = \begin{cases} 1, & \text{if } \sum_{c=1}^{C} d^{(c,p)} \geq \tau_C \\ 0, & \text{otherwise} \end{cases}$$
$$d^{(c,p)} = \begin{cases} 1, & l^{(c,p)} \geq \tau_s \\ 0, & \text{otherwise} \end{cases}, \tag{22}$$

where τ_C is the minimum number of channels to accept test epochs, and τ_s is a threshold for deciding whether to accept or reject the epochs for individual channels.

Our contribution involves training an F-channel LSIM (λ^c_{LSIM}) for $\hat{o}^{(c,p)}$ instead of a standard HMM (λ^c_{HMM}) regarding EEG channel c. We treat every single feature in $\hat{o}^{(c,p)}$ as a channel of LSIMs, and LSIMs capture the dynamic and interaction between individual features. For example, we train 12-channel LSIMs for MFCC features and 14-channel LSIMs for SPEC features. Then, we calculate the log-likelihood $l^{(c,p)} = P(\hat{o}^{(c,p)}|\lambda^c_{LSIM})$ for each epoch, and the verification is performed using the same decision rule as in (22).

In order to simulate a realistic usage scenario, data acquired during one session is used for training, while data acquired at later sessions are used to test the verification performances [40]. Hence, we also train the subject models using the data from the first session and test them independently using data from the second and third sessions. The verification performance is measured by the area under the curve (AUC). Two τ_C and τ_s thresholds are also determined by the same way used in [40]. In addition, we use four states and three Gaussian components per channel, like [40], [42].

For existing HMM-based and proposed LSIM-based methods, TABLE IV and TABLE V show AUC values for each stimulus (columns) and EEG feature (rows). The AUC values of HMMs have been taken directly from [40]. Additionally, the average performance across all types of stimuli per feature of the EEG is computed and reported in the last column. LSIMs have superior AUC results than HMMs in two test sessions for almost all stimuli and EEG feature types (bold values in columns). In the second and third sessions, the AUC values show that the proposed LSIM-based method has a significant improvements of 4.5% and 9.1% over the HMM-based method (paired t-test with α=0.01). The AUC improvement is also about 6.8% statistically significant under all conditions (stimulus, feature type, and session). The proposed LSIM-based method also reduces the standard deviation of AUC values across stimuli, EEG features, and sessions. For all conditions, the standard deviation of AUC values for HMM-based method is 5.4%, while it decreases significantly to 3.3% for LSIM-based method. Thus, LSIMs not only improve the verification performance but also decrease the standard deviation across all conditions.

D. Analysis of Approximate Inference Bias

Finally, we empirically examine how the bias of the approximate inference is transmitted to estimated parameters. So, the proposed EM algorithm is re-implemented based on exact inference. Exponential computation of the exact inference only allows us to consider LSIMs with few channels. We evaluate the effect of replacing exact inference with approximate inference by using simulated data from a 4-channel CHMM and a 4-channel configuration of the EPFL EEG dataset [38]. EM algorithms with exact and approximate parameters are compared for convergence speed and log-likelihood. Fig.8 shows the exact log-likelihood curves for EM algorithms via approximate and exact inferences for simulated data and a subject of the EPFL EEG dataset. These results are selected to represent the overall patterns that emerge from the results of EM algorithms via approximate and exact inferences for various situations. As can be seen, the exact inference gives faster convergence than approximate inference. Both plots confirm that approximate inferences reach parameters with

TABLE IV
VERIFICATION AUC RESULTS FOR SESSION 2 WHEN THE SYSTEM IS TRAINED WITH DATA FROM SESSION 1

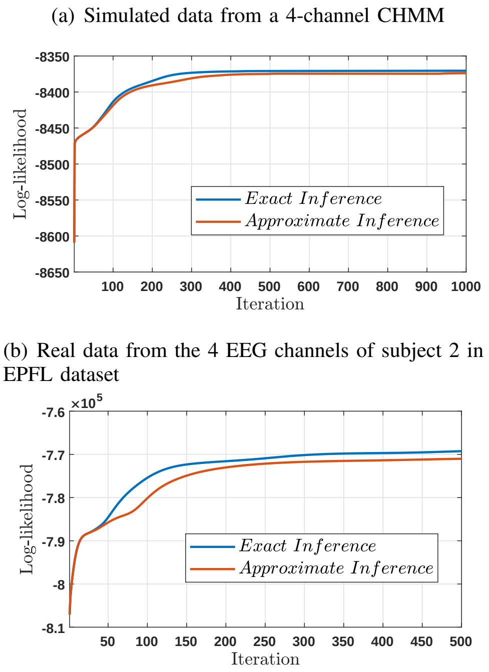
Model	Feature	AS	MC	RC	RO	VC3	VC5	VC7	VC10	VF3	VF5	VF7	VF10	Avg. (Std)
HMMs	MFCC	**77.0**	69.2	75.7	66.2	61.0	68.5	64.8	62.5	66.1	66.3	70.2	74.3	68.5 ± 4.9
	ARRC	73.2	70.3	72.8	67.7	54.2	66.1	67.2	57.2	64.3	68.1	74.8	74.8	68.5 ± 6.2
	SPEC	74.7	**70.4**	71.9	65.3	59.3	**75.7**	67.7	64.3	66.9	66.1	73.8	75.7	69.3 ± 5.0
LSIMs	MFCC	74.2	69.5	**75.8**	72.6	68.5	72.5	71.5	69.3	73.9	72.6	73.7	75.4	72.4 ± 2.4
	ARRC	70.9	69.0	73.1	**76.2**	**71.7**	73.3	**73.8**	68.6	75.1	**75.6**	74.7	**77.1**	73.2 ± 2.7
	SPEC	70.8	68.3	73.9	73.3	70.2	75.0	73.1	**73.7**	**75.5**	72.9	**75.1**	75.4	73.1 ± 2.3

Notes: Results in **bold** denote the best AUC per stimulus

TABLE V
VERIFICATION AUC RESULTS FOR SESSION 3 WHEN THE SYSTEM IS TRAINED WITH DATA FROM SESSION 1

Model	Feature	AS	MC	RC	RO	VC3	VC5	VC7	VC10	VF3	VF5	VF7	VF10	Avg. (Std)
HMMs	MFCC	71.1	67.1	70.3	73.2	65.1	70.7	65.2	55.4	60.5	75.8	68.7	64.9	67.3 ± 5.4
	ARRC	64.9	66.7	76.4	70.0	59.3	72.5	60.7	60.7	59.1	70.6	65.2	66.4	65.9 ± 5.2
	SPEC	63.8	62.1	69.2	71.2	63.2	69.6	65.7	55.7	63.6	67.6	68.1	62.6	65.2 ± 4.1
LSIMs	MFCC	**72.5**	**69.4**	78.5	77.3	75.6	78.7	**74.1**	69.7	73.3	78.9	76.7	73.4	74.8 ± 3.3
	ARRC	68.9	68.9	76.6	75.3	**78.2**	**82.5**	73.7	**76.8**	73.7	78.8	**77.1**	**77.7**	76.0 ± 3.9
	SPEC	68.4	68.7	**79.7**	**77.6**	76.6	80.7	73.8	71.8	**73.9**	**79.4**	75.2	74.4	75.1 ± 4.1

Notes: Results in **bold** denote the best AUC per stimulus


(a) Simulated data from a 4-channel CHMM

(b) Real data from the 4 EEG channels of subject 2 in EPFL dataset

Fig. 8. Log-likelihood convergence paths of exact and approximate EM algorithms (4-channel LSIM with five states and one Gaussian component)

nearly the same likelihood as exact inferences in the last iteration. The exact inference would give a better likelihood in some cases, but because of its computational complexity, it is not feasible to apply it to higher channel numbers.

V. CONCLUSION

We extend the scope of the re-estimation algorithm of HMMs for LSIMs in this study. Using the influence model and the multivariate mixture of strictly log-concave, we demonstrate that the LSIMs re-estimation converges to a local maximum of the likelihood function. The proposed auxiliary function has a unique global maximum, and closed-form re-estimation formulas are derived from marginal forward and backward parameters.

We test the theoretical convergence of the algorithm by examining simulated datasets and EEG/ECoG datasets (up to 64 channels). The log-likelihoods increase monotonically with iterations in all cases. As the model complexity increases, its log-likelihood increases as well, and it requires more iterations to reach convergence. Modeling and classification tasks compare the performance of LSIMs with standard HMMs. Modeling embedded Lorenz systems and ECoG recordings shows that LSIMs outperform HMMs according to AIC and BIC. A primary application of LSIMs as generative models is multi-channel time-series classification. CHMM data classification and EEG-based biometric verification are used to compare LSIMs and HMMs. The proposed LSIM-based method significantly improve the verification results over the existing HMM-based method, and it reduces the standard deviation of AUC values in all conditions. Therefore, LSIMs are suitable for modeling and classifying multi-channel time-series that exhibit spatial and temporal structure in many multi-channel signal processing applications. While exact inference gives a faster convergence rate than approximate inference, both inferences reach parameters with nearly the same likelihood in the last iteration.

This study and our previous study [28] solve both the inference and learning problems of LSIMs accurately and efficiently. However, there is an intrinsic difficulty in learning LSIMs and CHMMs due to having different channels with different states. In order to select the optimal state number per channel, the grid-search size increases exponentially as the channels increase. Future research can focus on optimizing search algorithms to deal with this exponential growth.

REFERENCES

[1] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, “Coupled hmm-based multi-sensor data fusion for sign language recognition,” *Pattern Recognition Letters*, vol. 86, pp. 1–8, 2017.

[2] A. V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy, "A coupled hmm for audio-visual speech recognition," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2, pp. II–2013, IEEE, 2002.

[3] A. H. Abdelaziz, S. Zeiler, and D. Kolossa, "Learning dynamic stream weights for coupled-hmm-based audio-visual speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 5, pp. 863–876, 2015.

[4] R. Zhao, G. Schalk, and Q. Ji, "Coupled hidden markov model for electrocorticographic signal classification," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pp. 1858–1862, IEEE, 2014.

[5] S. Zhong and J. Ghosh, "Hmms and coupled hmms for multi-channel eeg classification," in *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, vol. 2, pp. 1154–1159, IEEE, 2002.

[6] N. M. Ghahjaverestan, S. Masoudi, M. B. Shamsollahi, A. Beuchée, P. Pladys, D. Ge, and A. I. Hernández, "Coupled hidden markov model-based method for apnea bradycardia detection," *IEEE journal of biomedical and health informatics*, vol. 20, no. 2, pp. 527–538, 2016.

[7] T. Bolton and D. Van De Ville, "Sparse coupled hidden markov models shed light on resting-state fmri cross-network interactions," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pp. 358–361, Ieee, 2017.

[8] C. Sherlock, T. Xifara, S. Telfer, and M. Begon, "A coupled hidden markov model for disease interactions," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 62, no. 4, pp. 609–627, 2013.

[9] J. Kwon and K. Murphy, "Modeling freeway traffic with coupled hmms," tech. rep., Technical report, Univ. California, Berkeley, 2000.

[10] Y. Qi and S. Ishak, "A hidden markov model for short term prediction of traffic conditions on freeways," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 95 – 111, 2014. Special Issue on Short-term Traffic Flow Forecasting, issue no. 0968-090X.

[11] W. Cao, L. Cao, and Y. Song, "Coupled market behavior based financial crisis detection," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pp. 1–8, IEEE, 2013.

[12] K. Michalopoulos and N. Bourbakis, "Using dynamic bayesian networks for modeling eeg topographic sequences," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4928–4931, IEEE, 2014.

[13] G. Safont, A. Salazar, L. Vergara, E. Gómez, and V. Villanueva, "Multichannel dynamic modeling of non-gaussian mixtures," *Pattern Recognition*, vol. 93, pp. 312–323, 2019.

[14] J. Craley, E. Johnson, and A. Venkataraman, "A spatio-temporal model of seizure propagation in focal epilepsy," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1404–1418, 2019.

[15] M. Brand, "Coupled hidden markov models for modeling interacting processes," Tech. Rep. 405, MIT Media Lab Perceptual Computing/Learning and Common Sense, 1997.

[16] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden markov models for complex action recognition," in *Computer vision and pattern recognition, 1997. proceedings., 1997 ieee computer society conference on*, pp. 994–999, IEEE, 1997.

[17] S. Zhong and J. Ghosh, "A new formulation of coupled hidden markov models," tech. rep., Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, 2001.

[18] A. E. Raftery, "A model for high-order markov chains," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 47, no. 3, pp. 528–539, 1985.

[19] C. Asavathiratham and G. C. Verghese, *The influence model: A tractable representation for the dynamics of networked Markov chains*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2001.

[20] H. Ney, U. Essen, and R. Kneser, "On structuring probabilistic dependencies in stochastic language modelling," *Computer Speech & Language*, vol. 8, no. 1, pp. 1–38, 1994.

[21] L. K. Saul and M. I. Jordan, "Mixed memory markov models: Decomposing complex stochastic processes as mixtures of simpler ones," *Machine learning*, vol. 37, no. 1, pp. 75–87, 1999.

[22] W. Dong, "Influence modeling of complex stochastic processes," Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.

[23] W. Dong, B. Lepri, A. Cappelletti, A. S. Pentland, F. Pianesi, and M. Zancanaro, "Using the influence model to recognize functional roles in meetings," in *Proceedings of the 9th international conference on Multimodal interfaces*, pp. 271–278, ACM, 2007.

[24] W. Dong, A. Mani, A. Pentland, B. Lepri, and F. Pianesi, "Modeling group discussion dynamics," *IEEE Trans. Auton. Mental Dev*, submitted for publication.

[25] W. Pan, W. Dong, M. Cebrian, T. Kim, J. H. Fowler, and A. S. Pentland, "Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems," *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 77–86, 2012.

[26] W. Dong and A. Pentland, "Modeling influence between experts," in *Artifical Intelligence for Human Computing*, vol. 4451, pp. 170–189, Springer, 2007.

[27] W. Dong, *Modeling the structure of collective intelligence*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010.

[28] S. Karimi and M. B. Shamsollahi, "Tractable inference and observation likelihood evaluation in latent structure influence models," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5736–5745, 2020.

[29] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[31] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[32] B.-H. Juang, "Maximum-likelihood estimation for mixture multivariate stochastic observations of markov chains," *AT&T technical journal*, vol. 64, no. 6, pp. 1235–1249, 1985.

[33] L. Liporace, "Maximum likelihood estimation for multivariate observations of markov sources," *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 729–734, 1982.

[34] P. Touloupou, B. Finkenstädt, and S. E. Spencer, "Scalable bayesian inference for coupled hidden markov and semi-markov models," *Journal of Computational and Graphical Statistics*, vol. 29, no. 2, pp. 238–249, 2020.

[35] K. T. Abou-Moustafa, M. Cheriet, and C. Y. Suen, "Classification of time-series data using a generative/discriminative hybrid," in *Ninth International Workshop on Frontiers in Handwriting Recognition*, pp. 51–56, IEEE, 2004.

[36] K. Pillay, A. Dereymaeker, K. Jansen, G. Naulaers, S. Van Huffel, and M. De Vos, "Automated eeg sleep staging in the term-age baby using a generative modelling approach," *Journal of neural engineering*, vol. 15, no. 3, p. 036004, 2018.

[37] C. Sparrow, *The Lorenz equations: bifurcations, chaos, and strange attractors*, vol. 41. Springer Science & Business Media, 2012.

[38] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient p300-based brain–computer interface for disabled subjects," *Journal of Neuroscience methods*, vol. 167, no. 1, pp. 115–125, 2008.

[39] K. Shimoda, Y. Nagasaka, Z. C. Chao, and N. Fujii, "Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in japanese macaques," *Journal of neural engineering*, vol. 9, no. 3, p. 036015, 2012.

[40] P. Arnau-González, S. Katsigiannis, M. Arevalillo-Herráez, and N. Ramzan, "Bed: A new data set for eeg-based biometrics," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12219–12230, 2021.

[41] G. Celeux and G. Govaert, "Gaussian parsimonious clustering models," *Pattern recognition*, vol. 28, no. 5, pp. 781–793, 1995.

[42] E. Maiorana and P. Campisi, "Longitudinal evaluation of eeg-based biometric recognition," *IEEE transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1123–1138, 2017.