mode is a sample from the (high-dimensional) prior. Developing split-merge algorithms similar to those developed in [50] for the DP mixture model could be useful in ameliorating these issues.

Overall, the formulation we developed herein represents a flexible, Bayesian nonparametric model for describing complex dynamical phenomena and discovering simple underlying temporal structures.

Appendix

MNIW General Method: For the experiments of Section IV-A, we set $M = \mathbf{0}$ and $K = I_m$. This choice centers the mass of the prior around stable dynamic matrices while allowing for considerable variability. The inverse-Wishart portion is given $n_0 = m + 2$ degrees of freedom. For the HDP-AR-HMM, we set the scale matrix $S_0 = 0.75\bar{\Sigma}$, where $\bar{\Sigma} = 1/T \sum(\boldsymbol{y}_t - \bar{\boldsymbol{y}})(\boldsymbol{y}_t - \bar{\boldsymbol{y}})^T$. Setting the prior directly from the data can help move the mass of the distribution to reasonable values of the parameter space. Since each new considered dynamical mode is associated with a set of parameters sampled from the prior distribution and this dynamical mode is compared against others that have already been informed by the data, setting the base measure in this manner can improve mixing rates over a noninformative setting. For an HDP-SLDS with $\boldsymbol{x}_t \in \mathbb{R}^n$ and $\boldsymbol{y}_t \in \mathbb{R}^d$ and $n = d$, we set $S_0 = 0.675\bar{\Sigma}$. We then set the inverse-Wishart prior on the measurement noise, R, to have $r_0 = d + 2$ and $R_0 = 0.075\bar{\Sigma}$. For $n > d$, see [42].

Partially Supervised Honey Bee Experiments: For the partially supervised experiments of Section IV-C, we set $\Sigma_0 = 0.75S_0$. Since we are not shifting and scaling the observations, we set S_0 to 0.75 times the empirical covariance of the *first difference* observations. We also use $n_0 = 10$, making the distribution tighter than in the unsupervised case. Examining first differences is appropriate since the bee's dynamics are better approximated as a random walk than as i.i.d. observations. Using raw observations in the unsupervised approach creates a larger expected covariance matrix making the prior on the dynamic matrix less informative, which is useful in the absence of other labeled data.

Ibovespa Stock Index Experiments: For the HDP-SLDS variant of the MSSV model of (38), we rely on the N-IW-N prior described in Section V-A. For the dynamic parameter a and process noise mean $\mu^{(k)}$, we use $\mathcal{N}(0, 0.75\bar{\Sigma})$ priors. The IW prior on $\Sigma^{(k)}$ was given 3 degrees of freedom and an expected value of $0.75\bar{\Sigma}$. Finally, each component of the mixture-of-Gaussian measurement noise was given an IW prior with 3 degrees of freedom and an expected value of $5 * \pi^2$, which matches with the moment-matching technique of Harvey *et al.* [45].

For the HDP-SLDS comparison using the model of Table I, we use a MNIW prior with $M = 0$, $K = 1$, $n_0 = 3$ and $S_0 = 0.75\bar{\Sigma}$. The IW prior on R was given $r_0 = 100$ and an expected covariance of 25. Our sampler initializes parameters from the prior and we found it useful to set the prior around large values of R in order to avoid initial samples chattering between dynamical regimes caused by the state sequence having to account for the noise in the observations. After accounting for the residuals of the data in the posterior distribution, we typically learned $R \approx 10$.

References

[1] E. Fox, E. Sudderth, M. Jordan, and A. Willsky, "Nonparametric Bayesian learning of switching dynamical systems," *Adv. Neural Inf. Process. Syst.*, vol. 21, pp. 457–464, 2009.

[2] E. Fox, E. Sudderth, M. Jordan, and A. Willsky, "Nonparametric Bayesian identification of jump systems with sparse dependencies," in *Proc. 15th IFAC Symp. Syst. Identification*, Jul. 2009.

[3] V. Pavlovic, J. Rehg, and J. MacCormick, "Learning switching linear models of human motion," *Adv. Neural Inf. Process. Syst.*, vol. 13, pp. 981–987, 2001.

[4] L. Ren, A. Patrick, A. Efros, J. Hodgins, and J. Rehg, "A data-driven approach to quantifying natural human motion," *SIGGRAPH*, Aug. 2005.

[5] C.-J. Kim, "Dynamic linear models with Markov-switching," *J Econ.*, vol. 60, no. 1-2, pp. 1–22, Feb. 1994.

[6] M. So, K. Lam, and W. Li, "A stochastic volatility model with Markov switching," *J. Business Econom. Stat.*, vol. 16, no. 2, pp. 244–253, 1998.

[7] C. Carvalho and H. Lopes, "Simulation-based sequential analysis of Markov switching stochastic volatility models," *Comput. Statist. Data Anal.*, vol. 51, pp. 4526–4542, 2007.

[8] X. Li Rong and V. Jilkov, "Survey of maneuvering target tracking. Part V: Multiple-model methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, pp. 1255–1321, 2005.

[9] E. Fox, E. Sudderth, and A. Willsky, "Hierarchical Dirichlet processes for tracking maneuvering targets," in *Proc. Int. Conf. Inf. Fusion*, Jul. 2007.

[10] S. Oh, J. Rehg, T. Balch, and F. Dellaert, "Learning and inferring motion patterns using parametric segmental switching linear dynamic systems," *Int. J. Comput. Vis*, vol. 77, no. 1–3, pp. 103–124, 2008.

[11] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[12] V. Jilkov and X. Li Rong, "Online Bayesian estimation of transition probabilities for Markovian jump systems," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1620–1630, Jun. 2004.

[13] C. Li and S. Andersen, "Efficient blind system identification of non-Gaussian autoregressive models with HMM modeling of the excitation," *IEEE Trans. Signal Process.*, vol. 55, pp. 2432–2445, Jun. 2007.

[14] J. Chiang, Z. Wang, and M. McKeown, "A hidden Markov, multivariate autoregressive (HMM-MAR) network framework for analysis of surface EMG (SEMG) data," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4069–4081, Aug. 2008.

[15] X. Xuan and K. Murphy, "Modeling changing dependency structure in multivariate time series," in *Proc. Int. Conf. Mach. Learning*, Jun. 2007.

[16] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical Dirichlet processes," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 1, 2006.

[17] M. Beal, Z. Ghahramani, and C. Rasmussen, "The infinite hidden Markov model," *Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 577–584, 2002.

[18] J. Paisley and L. Carin, "Hidden Markov models with stick-breaking priors," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3905–3917, Oct. 2009.

[19] Y. Qi, J. Paisley, and L. Carin, "Music analysis using hidden Markov mixture models," *IEEE Trans. Signal Process.*, vol. 55, pp. 5209–5224, Nov. 2007.

[20] F. Caron, M. Davy, A. Doucet, E. Duflos, and P. Vanheeghe, "Bayesian inference for dynamic models with Dirichlet process mixtures," *IEEE Trans. Signal Process.*, vol. 56, pp. 71–84, Jan. 2008.

[21] E. Fox, E. Sudderth, M. Jordan, and A. Willsky, "A sticky HDP-HMM with application to speaker diarization," *Ann. Appl. Stat.*, 2011, to be published.

[22] D. MacKay, *Bayesian Methods For Backprop Networks*, ser. Models of Neural Networks, III. New York: Springer, 1994, ch. 6, pp. 211–254.

[23] *Bayesian Learning For Neural Networks, Ser. Lecture Notes in Statistics*, R. Neal, Ed. New York: Springer, 1996, vol. 118.

[24] M. Beal, "Variational algorithms for approximate Bayesian inference," Ph. D., University College London, London, U.K., 2003.

[25] S. Paoletti, A. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of hybrid systems: A tutorial," *Eur. J. Contr.*, vol. 2–3, pp. 242–260, 2007.

[26] R. Vidal, S. Soatto, Y. Ma, and S. Sastry, "An algebraic geometric approach to the identification of a class of linear hybrid systems," in *Proc. IEEE Conf. Decision Contr.*, Dec. 2003.

[27] Z. Psaradakis and N. Spagnolo, "Joint determination of the state dimension and autoregressive order for models with Markov regime switching," *J. Time Ser. Anal.*, vol. 27, pp. 753–766, 2006.