

# MIT 6.882 Final Project

Jeremiah Zhe Liu, Will Townes

May 1, 2016

## Instructions

The proposal (one per team) will be a written pdf document, 1-2 pages long, outlining the work to be done. It should include a plan with at least 4 steps, and indicate your internal deadlines for each of those steps.

If there are multiple participants, the division of responsibility should be made clear.

In addition, please include your assessment of what the risks to the project are: that is, what things do you think might turn out to be more difficult than planned, and what thoughts do you have about how to mitigate the risks?

If you are going to do an empirical study, be sure that you think about how you will evaluate your methodideally in comparison to other methods.

Remember that almost anything will turn out to be harder and more time-consuming than you expect. Try to arrange your project so that there are intermediate milestones that can serve as alternative finishing points, in case you dont get to the end. It will be much better to turn in a polished version of a small-scale project than to find yourself at the end of the term with a three-quarters implemented system of great depth and scope.

# 1 Introduction

The primary goal for our project is to gain experience with bayesian non-parametrics and time series modeling. We will approach this by attempting to replicate the paper “Nonparametric Bayesian Learning of Switching Linear Dynamical Systems” by Fox et al [1]. One motivation for learning these techniques comes from a biomedical application: the analysis of vital signs [2][3].

# 2 Plan and Deadlines

Key internal deadlines are listed in parentheses.

1. (3/21)  
Code and visualize a dirichlet process sampler. Visualize with a simple two-dimensional clustering simulation.
2. (3/28)  
Code and visualize a “sticky” hierarchical dirichlet process sampler. Visualize via a two-dimensional clusters-within-clusters simulation.
3. (4/4)  
Write a sampler for an ordinary (non-switching) linear dynamical system and test with simulation.
4. (4/11)  
Write a sampler for a simple hidden markov model with a fixed number of modes.
5. (4/18)  
Combine it all together with a switching linear dynamical system with sticky HDP prior. Use the conjugate prior specification.
6. (4/25)  
Write a sampler for a switching linear dynamical system with sticky HDP prior. Use the automatic relevance determination prior.
7. (5/2)  
Prepare draft of final report.

Additionally, if time permitting:

1. Apply the results to biomedical data and compare with [3]
2. Apply stochastic MCMC with the “FireflyMC” method [4] if possible.

### 3 Division of Responsibility

Each team member will attempt to write all of the samplers separately. We will compare results as we go along to help each other avoid getting stuck or making numerical errors. We will post our code in a public repository accessible at <https://github.com/willtownes/mit6882>. We may divide up the responsibilities for writing up the final reports. For example, one person might do more of the writing and the other person might spend more time producing graphics.

### 4 Risks

The primary risk of the project is getting distracted by responsibilities for other classes. We plan to mitigate this risk by striving to meet the internal deadlines and by holding each other accountable for making progress.

Another risk we face is that we do not have significant previous exposure to time series modeling or bayesian nonparametrics. However, we believe the former is not a major issue if we are following a bayesian approach, and the latter will be remedied in the upcoming class readings.

From a technical perspective, one difficulty could arise in slow mixing MCMC chains. We plan to mitigate this by working with low dimensional, small, simulated datasets until we are confident our methods are working.

Finally, it is possible our method may appear to work on simple datasets but fail on more realistic data. If we get to that point, we will try to apply our methods to the same datasets used in [1] (such as the honeybee waggle data) so that we can directly compare our results to the previous study.

## References

- [1] Emily Fox, Erik B. Sudderth, Michael I. Jordan, and Alan S. Willsky. Nonparametric Bayesian Learning of Switching Linear Dynamical Systems. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 457–464. Curran Associates, Inc., 2009.
- [2] L. H. Lehman, M. J. Johnson, S. Nemati, R. P. Adams, and R. G. Mark. Bayesian nonparametric learning of switching dynamics in cohort physiological time series: application in critical care patient monitoring. In *Advanced State Space Methods for Neural and Clinical Data*. Cambridge University Press, 2015.
- [3] L. w H. Lehman, R. P. Adams, L. Mayaud, G. B. Moody, A. Malhotra, R. G. Mark, and S. Nemati. A Physiological Time Series Dynamics-Based Approach to Patient Monitoring and Outcome Prediction. *IEEE Journal of Biomedical and Health Informatics*, 19(3):1068–1076, May 2015.
- [4] Dougal Maclaurin and Ryan P. Adams. Firefly Monte Carlo: Exact MCMC with Subsets of Data. *arXiv:1403.5693 [cs, stat]*, March 2014. arXiv: 1403.5693.