Nonparametric Bayesian Estimation of Switching Linear Dynamical Systems

Will Townes, Jeremiah Zhe Liu

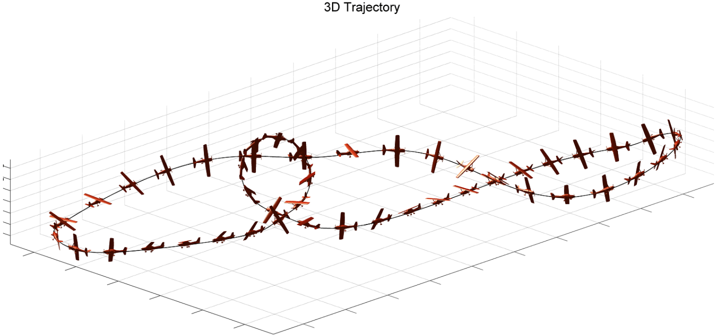
May 10, 2016

Overview

How to Estimate?

Fighter pilot's

- Maneuvering style: How was each maneuver executed?
- Maneuvering strategy: Maneuver combo's?

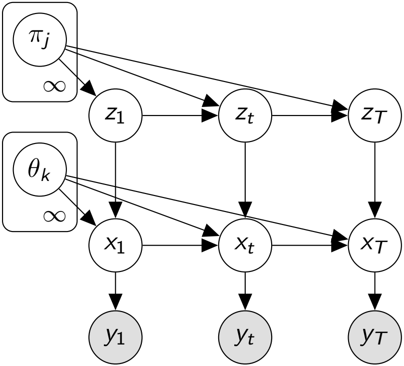


Switching Linear Dynamical System

$$\{\mathbf{A}, \mathbf{B}\} \in \mathcal{A} \times \mathcal{B} = \mathbf{\Theta}$$

$$z_t \sim Markov(\mathbf{\Pi})$$
$$x_t | z_t = \mathbf{A}_{z_t} x_{t-1} + \mathbf{B}_{z_t}$$
$$y_t | x_t \sim F(x_t)$$

Dimension?



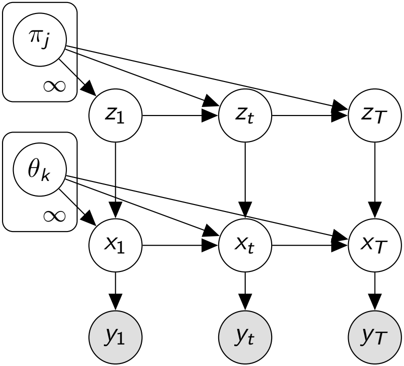
Switching Linear Dynamical System

$$\{\mathbf{A}, \mathbf{B}\} \in \mathcal{A} \times \mathcal{B} = \mathbf{\Theta}$$

$$z_t \sim Markov(\mathbf{\Pi})$$
$$x_t | z_t = \mathbf{A}_{z_t} x_{t-1} + \mathbf{B}_{z_t}$$
$$y_t | x_t \sim F(x_t)$$

Dimension?

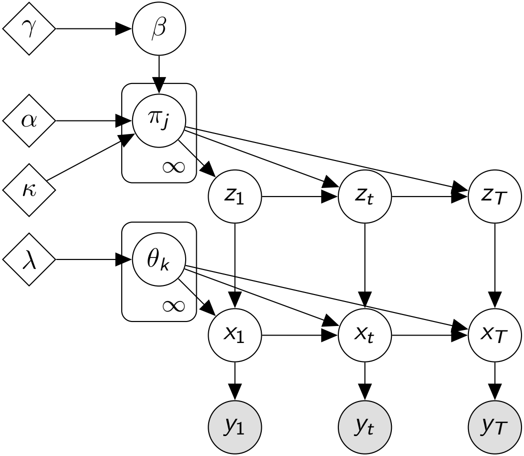
1. $dim(\mathcal{Z}) = |\mathcal{Z}|$
2. $dim(\mathbf{\Pi}) = O(|\mathcal{Z}|^2)$
3. $dim(\mathbf{\Theta}) = O(d_x^2 * |\mathcal{Z}|)$



Keep Calm...

and put a prior...

- $\pi_j \sim DP(\alpha, \beta) + \kappa\delta_j$
- But....

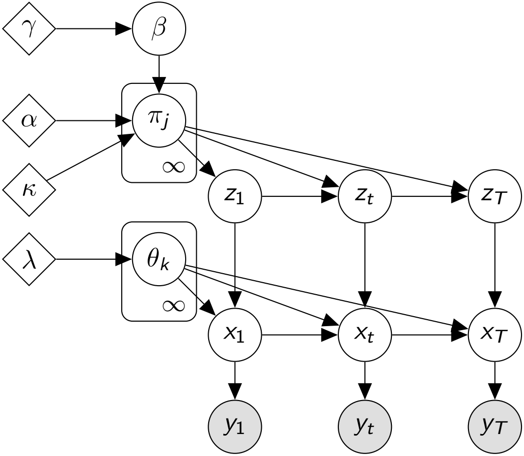


What does HDP mean in HMM?

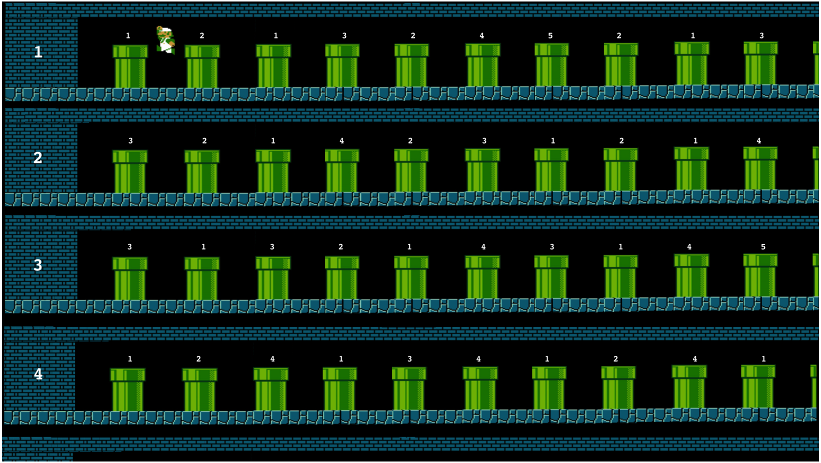
Restaurant Analogy?

- Global Dish Menu?
- Local Dish Menu?
- Restaurant?
- Dish?
- Customer?

Better Analogy?



HDP in HMM: Mario's Warp Pipe Process



HDP in SLDS: Inference Overview

Need to sample $(\mathbf{z}, \beta, \mathbf{\Pi})$ and $(\mathbf{x}, \mathbf{\Theta})$ jointly:

- Sample $\mathbf{z}, \beta, \mathbf{\Pi} | \mathbf{\Theta}, \mathbf{x}$
  1. Sample $\mathbf{z} \sim p(\mathbf{z}|\mathbf{\Pi}, \mathbf{\Theta}, \mathbf{x})$
     - In contrast to conditional sampler in original HDP $p(z_t = k|\mathbf{z}_{(-t)}, \beta) = p(z_t = k|\mathbf{y}) f(y_t|\mathbf{y}_{-(t)})$
     - efficiently algorithm available for Markov Model: Forward-backward Message Passing, i.e.

$$\mathbf{z} \sim p(\mathbf{z}|\mathbf{y}) = p(z_t|z_{t-1}, \mathbf{y}) p(z_{t-1}|z_{t-2}, \mathbf{y}) \dots p(z_2|z_1, \mathbf{y}) p(z_1|\mathbf{y})$$

  2. Sample β, π_k through standard CRF:
     - $\beta|\mathbf{z} \sim Dir(\mathbf{m}, \gamma)$
     - $\pi_k|\beta, \mathbf{z} \sim Dir(\alpha\beta_k + n_k)$
- Sample $\mathbf{x}, \mathbf{\Theta}|\mathbf{z}, \beta, \mathbf{\Pi}$

Linear Dynamical Systems

For times $t = 1 \dots T$, we are given data $y_t \in \mathbb{R}^n$. We assume y_t is a noise observation of a hidden, continuous state $x_t \in \mathbb{R}^d$, with $d \geq n$, and that x_t is a markov chain. The probability model is:

$$x_t | x_{t-1} \sim \mathcal{N}(Ax_{t-1} + B, \Sigma)$$
$$y_t | x_t \sim \mathcal{N}(Cx_t, R)$$

We assume C is known.

Switching Linear Dynamical Systems

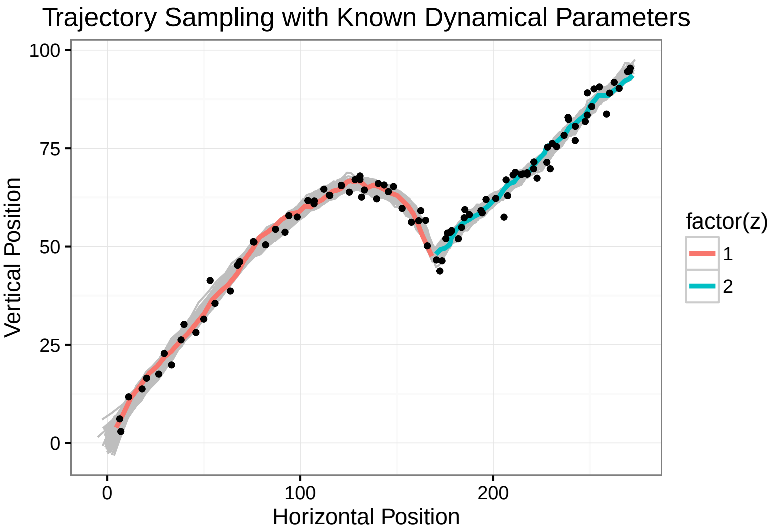
We now assume at each time t there is a hidden mode indexed by $z_t \in \{1, \ldots, K\}$ that determines the dynamical regime. The probability model is:

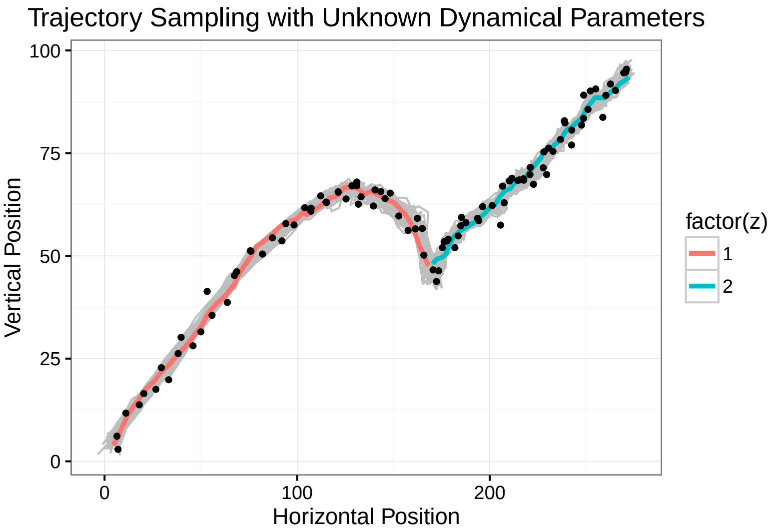
$$
\begin{aligned}
x_t | x_{t-1}, z_t &\sim \mathcal{N}(A^{(z_t)} x_{t-1} + B^{(z_t)}, \Sigma^{(z_t)}) \\
y_t | x_t &\sim \mathcal{N}(C x_t, R)
\end{aligned}
$$

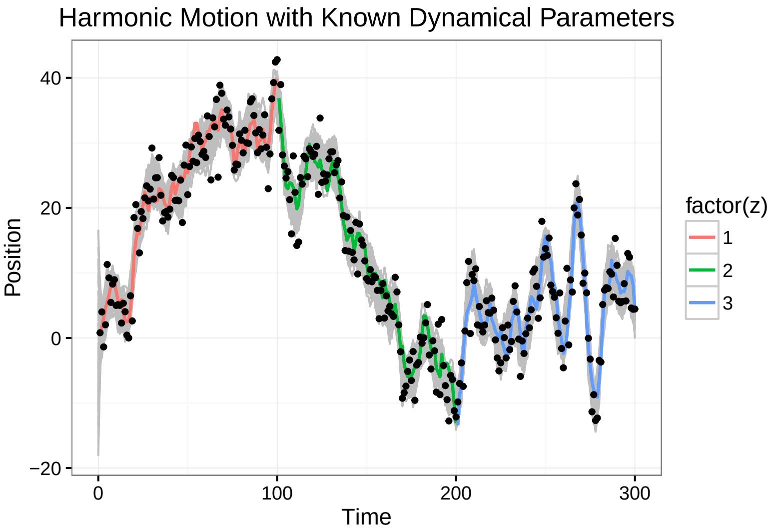
In the standard SLDS, $K < \infty$. In the HDP-HMM-SLDS, $K = \infty$.

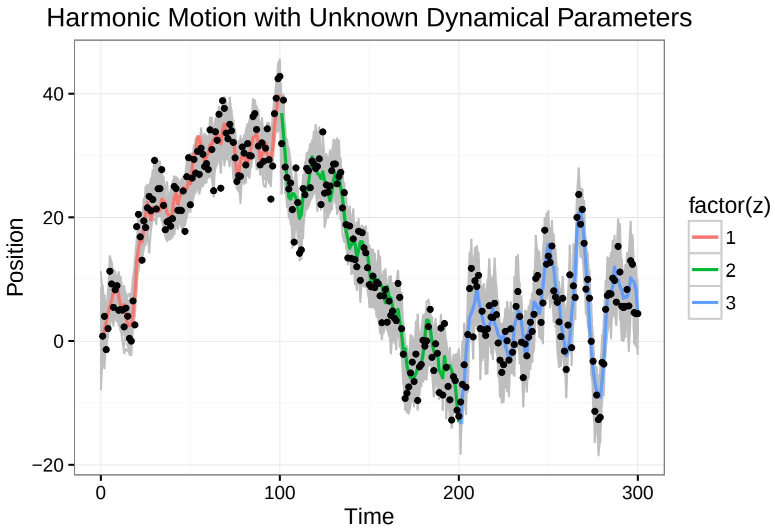
Algorithms for Inference

- Conditional on known dynamical parameters A, B, Σ and $z_{1:T}$, the hidden states $x_{1:T}$ are obtained via a **Kalman sampler**.
- Conditional on known hidden states $x_{1:T}$, the hidden modes $z_{1:T}$ are obtained via the **Forward-Backward** (sampling) algorithm.
- Conditional on known $x_{1:T}, z_{1:T}$, the dynamical parameters are obtained via **multivariate linear regressions** for each of the modes.


Trajectory Sampling with Known Dynamical Parameters
Vertical Position
Horizontal Position
0
25
50
75
100
0
100
200
factor(z)
1
2


Trajectory Sampling with Unknown Dynamical Parameters
Vertical Position
100
75
50
25
0
0
100
200
Horizontal Position
factor(z)
1
2


Harmonic Motion with Known Dynamical Parameters
Position
Time
40
20
0
−20
0
100
200
300
factor(z)
1
2
3


Harmonic Motion with Unknown Dynamical Parameters
Position
Time
40
20
0
−20
0
100
200
300
factor(z)
1
2
3

Lessons Learned

1. Kalman sampler great when dynamical parameters known.
2. When dynamical parameters unknown, choice of prior is crucial.
3. "Uninformative" hyperparameters didn't work. Need regularization.
4. Non-stationary time series present problems for empirical bayes.
5. Higher dimensional state space is more flexible but may overfit.