Scalable Bayesian reduced-order models for simulating high-dimensional multiscale dynamical systems.

P.S. Koutsourelakis, E. Bilionis

Cornell University
pk285@cornell.edu

ACDL Seminar Series
MIT April, 16 2010
Figure: Balloon floating in the wind. *An EM algorithm for Identification of Nonlinear Dynamical Systems* Sam Roweis & Zoubin Ghahramani, NIPS 98
In modern computational physics applications, we frequently need to solve high-dimensional deterministic (or stochastic) dynamical systems

\[
\frac{dy_t}{dt} = f(y_t) \quad \left( \frac{dy_t}{dt} = f(y_t; \omega) \right)
\]

- These arise from:
  - Discretization of transient PDEs
  - Inherently discrete dynamical systems (e.g. atomistic simulation)

In many other problems we have high-dimensional observables \(y_t\), but we do not have a model for the dynamics (e.g. meteorology)
In modern computational physics applications, we frequently need to solve high-dimensional deterministic (or stochastic) dynamical systems

$$\frac{dy_t}{dt} = f(y_t) \quad \left( \frac{dy_t}{dt} = f(y_t; \omega) \right)$$

- These arise from:
  - Discretization of transient PDEs
  - Inherently discrete dynamical systems (e.g. atomistic simulation)
- In many other problems we have high-dimensional observables $y_t$, but we do not have a model for the dynamics (e.g. meteorology)
Identify a reduced set of variables (reaction coordinates) \( x_t = P(y_t) \) \((\text{dim}(x_t) \ll \text{dim}(y_t))\) and derive a closed dynamical model:

\[
\frac{dx_t}{dt} = g(x_t), \quad x_t \in \hat{Y}
\]

- \( x_t \) are generally physically inspired, and in such cases knowing \( x_t \) usually gives us a pretty good idea of \( y_t = P^{-1}(x_t) \).
Identify a reduced set of variables (reaction coordinates) \( x_t = \mathcal{P}(y_t) \) (\( \text{dim}(x_t) \ll \text{dim}(y_t) \)) and derive a closed dynamical model:

\[
\frac{d x_t}{d t} = g(x_t), \quad x_t \in \hat{Y}
\]

More formal mathematical derivations are based on the Mori-Zwanzig framework (Zwanzig 1961) that produces a generalized Langevin equation for \( x_t \) (Chorin et al. 2000, PNAS, Darve et al. 2009, PNAS).
Motivation

Identify a reduced set of variables (reaction coordinates) \( x_t = \mathcal{P}(y_t) \) \((\text{dim}(x_t) \ll \text{dim}(y_t))\) and derive a closed dynamical model:

\[
\frac{dx_t}{dt} = g(x_t), \quad x_t \in \hat{Y}
\]

More formal mathematical derivations are based on the Mori-Zwanzig framework (Zwanzig 1961) that produces a generalized Langevin equation for \( x_t \) (Chorin et al. 2000, PNAS, Darve et al. 2009, PNAS).
Motivation

Identify a reduced set of variables (reaction coordinates) \( x_t = P(y_t) \) (\( \dim(x_t) \ll \dim(y_t) \)) and derive a closed dynamical model:

\[
\frac{dx_t}{dt} = g(x_t), \quad x_t \in \hat{Y}
\]

- Modern *concurrent* multiscale simulation environments such as the Equation-free method (Kevrekidis et al. 2004) or Heterogeneous Multiscale Method, HeMM (E. et al. 2003) need:
  - fine-scale model
  - *(coarse-scale model)*

  - restriction/compression (i.e. a mapping from fine \( y_t \) to coarse \( x_t \))
  - lifting/reconstruction (i.e. a mapping from coarse \( x_t \) to fine \( y_t \))
Motivation

Identify a reduced set of variables (reaction coordinates) \( x_t = \mathcal{P}(y_t) \) \( (\text{dim}(x_t) << \text{dim}(y_t)) \) and derive a closed dynamical model:

\[
\frac{dx_t}{dt} = g(x_t), \quad x_t \in \hat{y}
\]

Modern \textit{concurrent} multiscale simulation environments such as the Equation-free method (Kevrekidis et al. 2004) or Heterogeneous Multiscale Method, HeMM (E. at al. 2003) need:

- fine-scale model
- \textit{(coarse-scale model)}
- restriction/compression (i.e. a mapping from fine \( y_t \) to coarse \( x_t \))
- lifting/reconstruction (i.e. a mapping from coarse \( x_t \) to fine \( y_t \))
Objectives

- **Identification**: Can we identify those essential reaction coordinates $x_t$?
- **Identification**: Can we learn how the dynamics with respect to $x_t$ look like?
- **Prediction**: Can we evolve the lower-dimensional system forward in time and still predict where our original system will be in the long-run?
- **Uncertainty Quantification**: Can we quantify associated uncertainties that are inherent or simply arise from this reduced representation?

Desiderata:
- Model-Dimensionality Reduction
- Identification
- Prediction

pk285@cornell.edu
Cornell University

Bayesian reduced-order models for multiscale dynamical systems
Objectives

- **Identification**: Can we identify those essential reaction coordinates $x_t$?
- **Identification**: Can we learn how the dynamics with respect to $x_t$ look like?
- **Prediction**: Can we evolve the lower-dimensional system forward in time and still predict where our original system will be in the long-run?
- **Uncertainty Quantification**: Can we quantify associated uncertainties that are inherent or simply arise from this reduced representation?

**Desiderata**:
- Model-Dimensionality Reduction
- Identification
- Prediction
Existing Approaches - PCA or POD \(^1\)

- Assume \(y_t\) is stationary
- Evaluate sample covariance \(C = YY^T\), diagonalize \(C = PP^T\)
- Truncate after first few eigenvectors and approximate \(y_t = Px_t\). \((x_t \in \mathbb{R}^m, m << n)\)
- Determine dynamics in \(x\)-space:

\[
\frac{dx_t}{dt} = P^T f(Px)
\]

\(^1\)Pearson 1901, Lumley 1967

Figure: Observables
\[
Y = y_{t_i} = (y^{(1)}_{t_i}, y^{(2)}_{t_i})
\]
Motivation

- Most data-driven, dimensionality reduction techniques are for static data.
- Even though the reduced coordinates are learned from finite amounts of data, there is no quantification of the uncertainty associated with these inferences.
- Frequently, the dynamical systems we would be interested in coarse-graining are inherently stochastic.

We advocate probabilistic coarse-grained models that attempt to achieve dimensionality reduction and learn the reduced-dynamics simultaneously.
Motivation

- Most data-driven, dimensionality reduction techniques are for \textit{static data}.
- Even though the reduced coordinates are learned from finite amounts of data, there is no quantification of the uncertainty associated with these inferences.
- Frequently, the dynamical systems we would be interested in coarse-graining are inherently stochastic.

We advocate \textit{probabilistic coarse-grained models} that attempt to achieve dimensionality reduction and learn the reduced-dynamics \textit{simultaneously}.
Probabilistic State-Space models

These consist of:

- an unobserved, *latent* process $x_t$:
  
  coarse-grained dynamics: \( \frac{dx_t}{dt} = g(x_t; w_t) \) \((w_t : noise)\)

- the observed output $y_t$ arise as:
  
  fine-scale dynamics: \( y_t = h(x_t; u_t) \) \((u_t : noise)\)

- In most applications of such models $dim(y_t) \ll dim(x_t)$.
  For coarse-graining purposes though $dim(y_t) \gg dim(x_t)$

- $g$: describes the reduced dynamics

- $h$: describes the mapping from coarse to fine (i.e. the lifting/compression in EF or HeMM).
Probabilistic State-Space models

These consist of:

- an unobserved, latent process $x_t$:
  
  coarse-grained dynamics: $\frac{dx_t}{dt} = g(x_t; w_t)$ ($w_t : noise$)

- the observed output $y_t$ arise as:
  
  fine-scale dynamics: $y_t = h(x_t; u_t)$ ($u_t : noise$)

- Bayesian interpretation:
  
  - the first equation defines a prior probability distribution on $x_t$-path-space,
  
  - the second equation defines a likelihood
Probabilistic State-Space models

These consist of:

- an unobserved, \emph{latent} process $x_t$:

  coarse-grained dynamics: \[ \frac{dx_t}{dt} = g(x_t; w_t) \quad (w_t : \text{noise}) \]

- the observed output $y_t$ arise as:

  fine-scale dynamics: \[ y_t = h(x_t; u_t) \quad (u_t : \text{noise}) \]

Given $g$ and $h$, this is a basic an inverse problem, i.e. can you find $x_t$ that generated the data you observed?

This can be solved using your favorite filtering/smoothing algorithm (Kalman filter and its variation, particles filters etc)
Probabilistic State-Space models

These consist of:

- an unobserved, latent process $x_t$:
  
  coarse-grained dynamics: $\frac{dx_t}{dt} = g(x_t; w_t)$ \hspace{10pt} (w_t : noise)

- the observed output $y_t$ arise as:
  
  fine-scale dynamics: $y_t = h(x_t; u_t)$ \hspace{10pt} (u_t : noise)

- In our case though we do not know $g$ nor $h$
Probabilistic State-Space models

These consist of:

- an unobserved, *latent* process $x_t$:
  
  coarse-grained dynamics: $\frac{dx_t}{dt} = g(x_t; w_t)$ \quad ($w_t : noise$)

- the observed output $y_t$ arise as:
  
  fine-scale dynamics: $y_t = h(x_t; u_t)$ \quad ($u_t : noise$)

- In our case though we do not know $g$ nor $h$.
- Why do we need to know $h$?
Probabilistic State-Space models

These consist of:

- an unobserved, *latent* process $x_t$:
  
  coarse-grained dynamics: $\frac{dx_t}{dt} = g(x_t; w_t)$ \hspace{1cm} ($w_t : noise$)

- the observed output $y_t$ arise as:
  
  fine-scale dynamics: $y_t = h(x_t; u_t)$ \hspace{1cm} ($u_t : noise$)

In our case though we do not know $g$ nor $h$.

How can we learn $g$ nor $h$ in a manner that leads to sparse and interpretable coarse-grained models?
Switching Linear Dynamic System (SLDS) \(^2\)

**Motivation**

**Methodology**

Inference

**Numerical Examples**

---

Hidden State Process:

\[ S_t = 1, \ldots, K \]

\( K \) (hidden) SDEs in projected space, e.g.:

\[ \frac{d x^{(k)}_t}{dt} = b(x^{(k)}_t) + \sigma(x^{(k)}_t) dW^{(k)}_t, \quad k = 1, \ldots, K \]

Mixture model:

\[ y_t | S_t = k \sim N(P^{(k)} x^{(k)}_t, \Sigma^{(k)}) \]

---

\(^2\)Horenko et al. 2006

---

Figure: Observables

\[ Y = y_{t_i} = (y^{(1)}_{t_i}, y^{(2)}_{t_i}) \]
Switching Linear Dynamic System (SLDS)

**Figure:** Double-well potential for OU process

**Figure:** Time series for OU process with double-well potential

**Figure:** in the reduced-dimension space

Motivation

Methodology

Inference

Numerical Examples

pk285@cornell.edu

Cornell University

Bayesian reduced-order models for multiscale dynamical systems
Switching Linear Dynamic System (SLDS)

- Is it interpretable?

Figure: Approximation with two OU processes with harmonic potential

Figure: Time series for OU process with double-well potential
Switching Linear Dynamic System (SLDS)

**Motivation**

**Methodology**

**Inference**

**Numerical Examples**

Is it *sparse*?

**Figure:** Notably non-harmonic potential

**Figure:** Multimodal potential

pk285@cornell.edu

Cornell University

Bayesian reduced-order models for multiscale dynamical systems
Switching Linear Dynamic System (SLDS)

- This could still be expensive computationally if the number of hidden states needs to be large.
- Basic Assumption: At any point in time a single cluster (potential well) dictates the behavior of the system.
- Furthermore, there is no chance of identifying such a potential well unless it’s visited by the observables.
Switching Linear Dynamic System (SLDS)

- This could still be expensive computationally if the number of hidden states needs to be large.
- Basic Assumption: At any point in time a *single* cluster (potential well) dictates the behavior of the system.
- Furthermore, there is no chance of identifying such a potential well unless it’s visited by the observables.
Partial-Membership Linear Dynamic System (PMLDS)

- **Mixture of Experts** (Jacobs et al. 1997). For fixed $t$:

$$p(y \mid \Theta) = \sum_{k=1}^{K} \pi_k p_k(y \mid \theta_k)$$

can also be written as:

$$p(y \mid \Theta) = \sum_{z} p(z) \prod_{k=1}^{K} p_{z_k}^k(y \mid \theta_k)$$

where:

- $z = (z_1, \ldots, z_K)$, $z_k \in \{0, 1\}$ and $\sum_{k=1}^{K} z_k = 1$
- $p(z = (0, \ldots, 1, \ldots, 0)) = \pi_k$
Partial-Membership Linear Dynamic System (PMLDS)

- **Product of Experts** (Hinton 2002). What if we dropped the requirement that $z_k$ are binary and allowed them to take values in $[0, 1]$ such that $\sum_{k=1}^{K} z_k = 1$ still (Heller et al. 2008).

$$p(y | \Theta) = \sum p(z) \prod_{k=1}^{K} p_{z_k}^{z_k}(y_t | \theta_{k,t})$$
Partial-Membership Linear Dynamic System (PMLDS)

From a modeling perspective such an approach has several appealing properties:

- **(Sparseness)** Expressive ability does not hinge upon each individual component but rather is a result of its factorial character.

- **(Interpretability)** Intricate dynamical behavior can be captured and decomposed in terms of simple building blocks.

- It is highly-suited for problems that lack scale separation and where the evolution of the system is the result of phenomena at a cascade of scales.
Partial-Membership Linear Dynamic System (PMLDS)

- **Prior Specification:**
  - Dynamic membership $z_t = (z_{1,t}, \ldots, z_{k,t})$:
    - Let $\hat{z}_t$ a $K$ – dimensional isotropic OU process:
      \[
      d\hat{z}_t = -b_z (\hat{z}_t - q_z) + S_z dW_t
      \]
    - and (logistic normal):
      \[
      z_{k,t} = \frac{e^{\hat{z}_{k,t}} + 1/K}{\sum_{j=1}^{K} e^{\hat{z}_{j,t}} + 1}
      \]

- **Note:**
  - If $\hat{z}_{k,t} \rightarrow -\infty, \forall k$, then $z_{k,t} \rightarrow 1/K$.
  - If $\hat{z}_{k,t} \rightarrow +\infty, \forall k$, then $z_{k*,t} = 1$ where $k^*$ corresponds to max. $\hat{z}_{k,t}$.

Unknown parameters: $b_z, q_z, S_z$.

pk285@cornell.edu
Cornell University

Bayesian reduced-order models for multiscale dynamical systems
Partial-Membership Linear Dynamic System (PMLDS)

- **Prior Specification:**
  - Dynamic membership $z_t = (z_{1,t}, \ldots, z_{k,t})$:
    - Let $\hat{z}_t$ a $K$-dimensional isotropic OU process:
      \[
      d\hat{z}_t = -b_z (\hat{z}_t - q_z) + S_z dW_t
      \]
    - and (logistic normal):
      \[
      z_{k,t} = \frac{e^{\hat{z}_{k,t}} + 1/K}{\sum_{j=1}^{K} e^{\hat{z}_{j,t}} + 1}
      \]
  - Note:
    - If $\hat{z}_{k,t} \to -\infty, \forall k$, then $z_{k,t} \to \frac{1}{K}$.
    - If $\hat{z}_{k,t} \to +\infty, \forall k$, then $z_{k^*,t} = 1$ where $k^*$ corresponds to max. $\hat{z}_{k,t}$
    - **Unknown parameters:** $b_z, q_z, S_z$

pk285@cornell.edu
Cornell University

Bayesian reduced-order models for multiscale dynamical systems
Partial-Membership Linear Dynamic System (PMLDS)

- Prior Specification:
  - Latent dynamics $x_t^{(k)} \in \mathbb{R}^M$ ($M \ll \text{dim}(y_t)$) of each model $k$. For example an isotropic OU-process:
    \[
    dx_t^{(k)} = -b_x^{(k)} (x_t^{(k)} - q_x^{(k)}) + S^{(k)} dW_t^{(k)}
    \]

- Physical insight can be helpful here, but not necessary
- Unknown parameters: $\{b_x^{(k)}, q^{(k)}, S^{(k)}\}_{k=1}^{K}$
Partial-Membership Linear Dynamic System (PMLDS)

- Prior Specification:
  - Latent dynamics $x_t^{(k)} \in \mathbb{R}^M (M << \text{dim}(y_t))$ of each model $k$. For example an isotropic OU-process:
    \[
    dx_t^{(k)} = -b_x^{(k)} (x_t^{(k)} - q_x^{(k)}) + S^{(k)} dW_t^{(k)}
    \]
  - Physical insight can be helpful here, but not necessary
  - Unknown parameters: $\{b_x^{(k)}, q^{(k)}, S^{(k)}\}_{k=1}^K$
Partial-Membership Linear Dynamic System (PMLDS)

Bayesian model

\[
p(\Theta \mid \{y_t\}_{t=1}^T) \propto p(\{y_t\}_{t=1}^T \mid \Theta) p(\Theta)
\]

Posterior \quad Likelihood \quad Prior

Likelihood. Assume observations \(y_t \in \mathbb{R}^n\) are available at discrete, equidistant time instants \(t_i = i\tau\). Then:

\[
y_t \mid x_t, z_t \sim N(\mu_{y,t}, \Sigma)
\]

where:

- \(\mu_{y,t} = \sum_{k=1}^{K} z_{k,t} P^{(k)} x_t^{(k)}\)
- \(\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)\)
- Unknown parameters: \(\{P^{(k)}\}_{k=1}^K, \Sigma\)
Partial-Membership Linear Dynamic System (PMLDS)

Bayesian model

\[ p(\Theta | \{y_t\}_{t=1}^T) \propto p(\{y_t\}_{t=1}^T | \Theta) p(\Theta) \]

- **Likelihood.** Assume observations \( y_t \in \mathbb{R}^n \) are available at discrete, equidistant time instants \( t_i = i\tau \). Then:

\[
y_t \mid x_t, z_t \sim N(\mu_{y,t}, \Sigma)
\]

where:

- \( \mu_{y,t} = \sum_{k=1}^{K} z_{k,t} P^{(k)} x_t^{(k)} \)
- \( \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) \)
- **Unknown parameters:** \( \{P^{(k)}\}_{k=1}^{K}, \Sigma \)
Partial-Membership Linear Dynamic System (PMLDS)

Example:

Figure: Realizations of two hidden ($K = 2$) one-dimensional ($M = 1$) Ornstein-Uhlenbeck processes
Partial-Membership Linear Dynamic System (PMLDS)

Example:

Figure: Two metastable states
Partial-Membership Linear Dynamic System (PMLDS)

Example:

(a) observable

(b) histogram

Figure: *Three metastable states*
Partial-Membership Linear Dynamic System (PMLDS)

Example:

(a) observable

(b) histogram

Figure: One metastable state
Partial-Membership Linear Dynamic System (PMLDS)

- Given observations $y_{1:T}$, the model parameters are:
  - **Dynamic $\Theta_{1:T}$**:
    - memberships $z_{1:T}$, $(\hat{z}_{1:T})$
    - latent reduced-dynamics $\{x^{(k)}_{1:T}\}_{k=1}^K$
  - **Static $\Theta$**:
    - $b_z, q_z, S_z$ (prior for $z_t$)
    - $\{b_x^{(k)}, q^{(k)}, S^{(k)}\}_{k=1}^K$ (prior for $x^{(k)}_t$)
    - $\{P^{(k)}\}_{k=1}^K, \Sigma = diag(\sigma_1^2, \ldots, \sigma_n^2)$ (likelihood)

- Learning/Inference:
  - Priors can be placed on the static parameters as well
  - Note that some of these parameters (e.g., $P^{(k)}$) are of dimension $n = \dim(y_t) \gg 1$
  - Point estimates $\Theta^*$ (MLE or MAP) for static parameters and sampling from the posterior $p(\Theta_{1:T} | \Theta^*, y_{1:T})$
Partial-Membership Linear Dynamic System (PMLDS)

Given observations $y_{1:T}$, the model parameters are:

1. **Dynamic $\Theta_{1:T}$:**
   - memberships $z_{1:T}, (\hat{z}_{1:T})$
   - latent reduced-dynamics $\{x^{(k)}_{1:T}\}_{k=1}^K$

2. **Static $\Theta$:**
   - $b_z, q_z, S_z$ (prior for $z_t$)
   - $\{b^{(k)}_x, q^{(k)}_z, S^{(k)}\}_{k=1}^K$ (prior for $x^{(k)}_t$)
   - $\{P^{(k)}\}_{k=1}^K, \Sigma = \text{diag}(\sigma^2_1, \ldots, \sigma^2_n)$ (likelihood)

Learning/Inference:
- Priors can be placed on the static parameters as well.
- Note that some of these parameters (e.g., $P^{(k)}$) are of dimension $n = \dim(y_t) >> 1$.
- Point estimates $\Theta^*$ (MLE or MAP) for static parameters and sampling from the posterior $p(\Theta_{1:T} | \Theta^*, y_{1:T})$.
Partial-Membership Linear Dynamic System (PMLDS)

- Given observations $y_{1:T}$, the model parameters are:
  1. **Dynamic $\Theta_{1:T}$**:
     - memberships $z_{1:T}$, $(\hat{z}_{1:T})$
     - latent reduced-dynamics $\{x_{1:T}^{(k)}\}_{k=1}^K$
  2. **Static $\Theta$**:
     - $b_z, q_z, S_z$ (prior for $z_t$)
     - $\{b_x^{(k)}, q^{(k)}, S^{(k)}\}_{k=1}^K$ (prior for $x_t^{(k)}$)
     - $\{P^{(k)}\}_{k=1}^K, \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ (likelihood)

- Learning/Inference:
  - Priors can be placed on the static parameters as well
  - Note that some of these parameters (e.g. $P^{(k)}$) are of dimension $n = \text{dim}(y_t) >> 1$
  - point estimates $\Theta^*$ (MLE or MAP) for static parameters and sampling from the posterior $p(\Theta_{1:T} | \Theta^*, y_{1:T})$
Partial-Membership Linear Dynamic System (PMLDS)

Given observations $y_{1:T}$, the model parameters are:

1. **Dynamic $\Theta_{1:T}$**:
   - memberships $z_{1:T}, (\hat{z}_{1:T})$
   - latent reduced-dynamics $\{x_{1:T}^{(k)}\}_{k=1}^{K}$

2. **Static $\Theta$**:
   - $b_z, q_z, S_z$ (prior for $z_t$)
   - $\{b_{x}^{(k)}, q^{(k)}, S^{(k)}\}_{k=1}^{K}$ (prior for $x_t^{(k)}$)
   - $\{P^{(k)}\}_{k=1}^{K}, \Sigma = \text{diag}(\sigma^2_1, \ldots, \sigma^2_n)$ (likelihood)

**Learning/Inference**:
- Priors can be placed on the static parameters as well
- Note that some of these parameters (e.g. $P^{(k)}$) are of dimension $n = \text{dim}(y_t) >> 1$
- point estimates $\Theta^*$ (MLE or MAP) for static parameters and sampling from the posterior $p(\Theta_{1:T} \mid \Theta^*, y_{1:T})$
Partial-Membership Linear Dynamic System (PMLDS)

- Given observations $y_{1:T}$, the model parameters are:
  1. **Dynamic $\Theta_{1:T}$**:
     - memberships $z_{1:T}, (\hat{z}_{1:T})$
     - latent reduced-dynamics $\{x_{1:T}^{(k)}\}_{k=1}^{K}$
  2. **Static $\Theta$**:
     - $b_z, q_z, S_z$ (prior for $z_t$)
     - $\{b_x^{(k)}, q^{(k)}, S^{(k)}\}_{k=1}^{K}$ (prior for $x_t^{(k)}$)
     - $\{P^{(k)}\}_{k=1}^{K}, \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ (likelihood)

- Learning/Inference:
  - Priors can be placed on the static parameters as well
  - Note that some of these parameters (e.g. $P^{(k)}$) are of dimension $n = \text{dim}(y_t) >> 1$
  - point estimates $\Theta^*$ (MLE or MAP) for static parameters and sampling from the posterior $p(\Theta_{1:T} \mid \Theta^*, y_{1:T})$
Partial-Membership Linear Dynamic System (PMLDS)

Expectation-Maximization

E-step - Iteration $j$:

$$Q(\Theta, \Theta_j) = \int \log p_\Theta(q_{1:T}, y_{1:T}) p_{\Theta_j}(q_{1:T}, y_{1:T}) dq_{1:T}$$

M-step:

$$\Theta_{i+1} = \arg\max_{\Theta} Q(\Theta, \Theta_i)$$
Partial-Membership Linear Dynamic System (PMLDS)

Expectation-Maximization

- **E-step - Iteration** $j$:
  \[ Q(\Theta, \Theta_j) = \int \log p_\Theta(q_{1:T}, y_{1:T}) \, p_{\Theta_j}(q_{1:T}, y_{1:T}) \, dq_{1:T} \]

- **M-step**:
  \[ \Theta_{j+1} = \operatorname{argmax}_{\Theta} Q(\Theta, \Theta_j) \]

or with respect to sufficient statistics:

- **E-step - Iteration** $j$:
  \[ \Phi_{j+1} = E_{\Theta_j}[\Psi(q_{1:T}, y_{1:T})] \]

- **M-step**:
  \[ \Theta_{j+1} = \Lambda(\Phi_{j+1}) \text{ (i.e. invert the relation with } \Phi) \]
Partial-Membership Linear Dynamic System (PMLDS)

Expectation-Maximization

- **E-step - Iteration** \( j \):
  \[
  \Phi_j = E_{\Theta_j} \left[ \psi(q_{1:T}, y_{1:T}) \right]
  \]

- **M-step:**
  \[
  \Theta_{j+1} = \Lambda(\Phi_j) \text{ (i.e. invert the relation with } \Phi) \]

- A *recursive* (online) version is available which can employ SMC (particle filters) to perform the E-step.

- This essentially implies performing importance sampling in spaces of increasing dimension which generally fails due to error accumulation over time.
Partial-Membership Linear Dynamic System (PMLDS)

Split-data likelihood or Pseudo-likelihood (Online EM, Andrieu et al. 2005)

If we break the data $y_{1:T}$ into $B$ blocks of length $L$ we obtain a pseudo-log-likelihood:

$$I_L(\Theta, y_{1:BL}) = \sum_{b=1}^{B} \log p_\Theta(y_{(b-1)L+1:bL})$$

which ignores the dependence between blocks.

- It can be shown that the $\Theta$ maximizing the split-data pseudo-likelihood includes the true value.
- Maximization requires expectation over finite horizon of length $L$. 
Split-data likelihood or Pseudo-likelihood (Online EM, Andrieu et al. 2005)

If we break the data $\mathbf{y}_{1:T}$ into $B$ blocks of length $L$ we obtain a pseudo-log-likelihood:

$$I_L(\Theta, \mathbf{y}_{1:BL}) = \sum_{b=1}^{B} \log p_{\Theta}(\mathbf{y}_{(b-1)L+1:bL})$$

which ignores the dependence between blocks.

- It can be shown that the $\Theta$ maximizing the split-data pseudo-likelihood includes the true value.
- Maximization requires expectation over finite horizon of length $L$. 
Partial-Membership Linear Dynamic System (PMLDS)

Split-data likelihood or Pseudo-likelihood (Online EM, Andrieu et al. 2005)

If we break the data $\mathbf{y}_{1:T}$ into $B$ blocks of length $L$ we obtain a pseudo-log-likelihood:

$$I_L(\Theta, \mathbf{y}_{1:BL}) = \sum_{b=1}^{B} \log p_\Theta(\mathbf{y}_{(b-1)L+1:bL})$$

which ignores the dependence between blocks.

- It can be shown that the $\Theta$ maximizing the split-data pseudo-likelihood includes the true value.
- Maximization requires expectation over *finite horizon of length* $L$. 
Online EM

Let $\gamma_j$ such that $\sum_j \gamma_j \to +\infty, \sum_j \gamma_j^2 \to c < +\infty$. (e.g. $\gamma_j = 1/j$.

Then at each iteration $j$:

- **E-Step:**
  - Run SMC over horizon $L$ using stationary distribution and evaluate $E_{\Theta_j}[^\Psi(q_{(j-1)L+1:jL}, y_{(j-1)L+1:jL})]$.
  - Update sufficient statistics:
    \[
    \Phi_{j+1} = (1 - \gamma_j) + \gamma_j E_{\Theta_j}[^\Psi(q_{(j-1)L+1:jL}, y_{(j-1)L+1:jL})]
    \]

- **M-Step:**
  \[
  \Theta_{j+1} = \Lambda(\Phi_{j+1}) \text{ (i.e. invert the relation with } \Phi)\]
Partial-Membership Linear Dynamic System (PMLDS)

Some features of the algorithm

- operations are $O(n)$ (where $n \gg 1$ is the dimension of the observables)
- memory requirements are $O(n)$
- embarrassingly parallelizable
- update equations for static parameters $\Theta$ require a few fixed-point iterations at each step.
Example 1

- Data $y_t$: Average daily temperatures at 50 locations in US states from 1/1/1995 to 1/13/2009.
Example 1

- Data $y_t$: Average daily temperatures at 50 locations in US states from 1/1/1995 to 1/13/2009.
Example 1

Figure: (Top) Posterior mean of $z_{m,t}$, $m = 1, 2$ based on the SLDS and (Bottom) particulate approximation of the posterior of $z_{m,t}$, $m = 1, 2$ PMLDS.
Example 1

<table>
<thead>
<tr>
<th>$K$</th>
<th>$M$</th>
<th>SLDS</th>
<th>PMLDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>$-179.97 \pm 37.31$</td>
<td>$-171.11 \pm 37.20$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$-170.68 \pm 36.95$</td>
<td>$-141.11 \pm 27.82$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>$-176.40 \pm 34.36$</td>
<td>$-143.81 \pm 25.56$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>$-166.05 \pm 30.57$</td>
<td>$-117.67 \pm 21.15$</td>
</tr>
</tbody>
</table>

Table: One-step-ahead predictive log-likelihood of SLDS and PMLDS models for various $M$, $K$.

- $K$: number of experts
- $M$: dimension of the dynamics of each expert
Example 2

Multiscale PDEs:

\[
\begin{align*}
&u_t(x, t) = \frac{d}{dx} \left( a_\epsilon(x) \frac{du(x,t)}{dt} \right) \\
&x \in [0, 1], \ u(0, t) = u(1, t) = 0
\end{align*}
\]
Example 2

- Multiscale PDEs:
  \[
  \begin{align*}
  u_t(x, t) &= \frac{d}{dx} \left( a_\epsilon(x) \frac{du(x,t)}{dt} \right) \\
  x &\in [0, 1], \quad u(0, t) = u(1, t) = 0
  \end{align*}
  \]

- Discretization in space \(x\): \(\{u(x_i, t)\} = y_t\)
  \[
  \frac{dy_t}{dt} = A y_t
  \]

- Discretization in time \(t\) with time step \(\tau = 10^{-4}\)
Example 2

Temperature profile $u(x, T)$ at various $T$
Example 2

Temperature profile \( u(x, T) \) at various \( T \)
Example 2

Temperature profile $u(x, T)$ at various $T$

- $T=0$ (initial)
- $T=0.001$
- $T=0.01$
Example 2

Temperature profile $u(x, T)$ at various $T$
Example 2

Temperature profile $u(x, T)$ at various $T$
Example 2

Temperature profile $u(x, T)$ at various $T$
Example 2

**Figure:** *Bayesian adaptive time-integration*

Bayesian reduced-order models for multiscale dynamical systems
Example 2

**Prediction using PMLDS:**

1. 2 hidden states/modes.
2. Full system is solved for 20 time steps i.e. $\delta t = 2 \times 10^{-3}$.
3. Inferred PMLDS is used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$.
4. Full system is re-initialized at posterior mean and run for another 20 time steps i.e. $\delta t = 2 \times 10^{-3}$.
5. PMLDS is updated with new data and used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$.
6. Goto step 3

Computational acceleration by a factor of 25
Dimensionality reduction by a factor of 500
Example 2

- **Prediction using PMLDS:**
  1. 2 hidden states/modes.
  2. Full system is solved for 20 time steps i.e. $\delta t = 2 \times 10^{-3}$
  3. Inferred PMLDS is used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$
  4. Full system is re-initialized at posterior mean and run for another 20 time steps i.e. $\delta t = 2 \times 10^{-3}$
  5. PMLDS is updated with new data and used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$
  6. Goto step 3

Computational acceleration by a factor of 25
Dimensionality reduction by a factor of 500
Example 2

**Prediction using PMLDS:**

1. 2 hidden states/modes.
2. Full system is solved for 20 time steps i.e. $\delta t = 2 \times 10^{-3}$
3. Inferred PMLDS is used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$

4. Full system is re-initialized at posterior mean and run for another 20 time steps i.e. $\delta t = 2 \times 10^{-3}$.
5. PMLDS is updated with new data and used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$.
6. Goto step 3

---

Computational acceleration by a factor of 25
Dimensionality reduction by a factor of 500
Example 2

**Prediction using PMLDS:**

1. 2 hidden states/modes.
2. Full system is solved for 20 time steps i.e. $\delta t = 2 \times 10^{-3}$
3. Inferred PMLDS is used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$
4. Full system is re-initialized at *posterior mean* and run for another 20 time steps i.e. $\delta t = 2 \times 10^{-3}$.
5. PMLDS is updated with new data and used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$.
6. Goto step 3

Computational acceleration by a factor of 25
Dimensionality reduction by a factor of 500
Example 2

**Prediction using PMLDS:**

1. 2 hidden states/modes.
2. Full system is solved for 20 time steps i.e. $\delta t = 2 \times 10^{-3}$
3. Inferred PMLDS is used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$
4. Full system is re-initialized at posterior mean and run for another 20 time steps i.e. $\delta t = 2 \times 10^{-3}$.
5. PMLDS is updated with new data and used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$.
6. Goto step 3

Computational acceleration by a factor of 25
Dimensionality reduction by a factor of 500

pk285@cornell.edu
Cornell University
Bayesian reduced-order models for multiscale dynamical systems
Example 2

**Prediction using PMLDS:**

1. 2 hidden states/modes.
2. Full system is solved for 20 time steps i.e. $\delta t = 2 \times 10^{-3}$
3. Inferred PMLDS is used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$
4. Full system is re-initialized at *posterior mean* and run for another 20 time steps i.e. $\delta t = 2 \times 10^{-3}$.
5. PMLDS is updated with new data and used to “evolve” the system for 500 time steps, i.e. $\delta T = 5 \times 10^{-2}$.
6. Goto step 3

**Computational acceleration by a factor of 25**

**Dimensionality reduction by a factor of 500**
Figure: Comparison of predictive posterior estimates (posterior mean and 5% and 95% quantiles) with exact solution $u(x, t)$ at $t = 0.15$
**Example**

*Figure:* Comparison of predictive posterior estimates (posterior mean and 5% and 95% quantiles) with exact solution $u(x, t)$ at $t = 0.25$
Figure: *Comparison of predictive posterior estimates (posterior mean and 5% and 95% quantiles) with exact solution $u(x, t)$ at $t = 0.5*
Example

Figure: Comparison of predictive posterior estimates (posterior mean and 5% and 95% quantiles) with exact solution $u(x, t)$ at $t = 1.0$
Example 2

How would the results look if we wanted a higher acceleration (i.e. increased $\delta T/\delta t$)?

Figure: *Comparison at $t = 0.25$*
Example 2

How would the results look if we wanted a higher acceleration (i.e. increased $\delta T/\delta t$)?

Figure: Comparison at $t = 1.0$
Example 2

How would the results look if we wanted a higher acceleration (i.e. increased $\delta T/\delta t$)?

![Graphs showing comparison at $t = 1.0$](image)

**Figure:** *Comparison at $t = 1.0$*
Conclusions

- A lightweight, structure prior model has been proposed for developing reduced-order models for high-dimensional dynamical systems.
- It employs a SMC scheme for the dynamic hidden states and MAP (or MLE) for the static parameters.
- Additional data can be ingested *sequentially*.
- The inferred model can be readily used to predict fast and accurately the evolution of the original system in time.
- In multiscale computations, the predictive posterior can be used to adaptively determine when additional data from computations are needed.