Data-driven model reduction using sparse Bayesian learning: Application to nonlinear aeroelastic system

Rimple Sandhu¹, Mohammad Khalil², Chris Pettit³, Dominique Poirel⁴ and Abhijit Sarkar¹

¹Carleton University, Ottawa, ON, Canada
²Sandia National Lab, Livermore, CA, USA
³United States Naval Academy, Annapolis, MA, USA
⁴Royal Military College of Canada, Kingston, ON, Canada

Research Challenges and Opportunities at the interface of Machine Learning and Uncertainty Quantification
University of Southern California
June 4-6, 2018
Overview

Context of this talk
Bayesian inversion methodology
Sparse bayesian learning (SBL)
SBL for nonlinear dynamical systems
Application: Nonlinear aeroelastic oscillator
Conclusion and future direction
Context of this talk

Bayesian inversion methodology
Sparse bayesian learning (SBL)
SBL for nonlinear dynamical systems
Application: Nonlinear aeroelastic oscillator
Conclusion and future direction
Context of the talk: Inverse modelling of nonlinear dynamical systems

What we have?
- Noisy, sparse and incomplete observations from the system
- Limited understanding/literature of the underlying physics
- Multiple mathematical models exist that can fit the observations reasonably well

What we want?
- Build a mathematical model that respects modelling and measurement uncertainties
- Robust prediction capability for a quantity-of-interest
- Identify and quantify the nonlinear physics for control, sensitivity or design considerations

Solution? Bayesian inversion methodology:
- Stochastic, nonlinear state-space model structure
- Bayesian state/parameter estimation
- Bayesian model selection/averaging [Focus of this talk]
- Posterior predictive distribution

Practical issues with Bayesian model selection for nonlinear dynamical systems?
- How to assign parameter prior pdfs to model parameters with limited knowledge? Will flat priors work?
- Which models to consider for comparison? Are we missing out on better models?

Sparse Bayesian learning (SBL)/Automatic relevance determination (ARD)
Bayesian inversion methodology
Bayesian inversion methodology

- General discrete-time, stochastic, state-space representation for each model $M_i \in M$:

  \[
  \text{Model equation: } u_{k+1} = g_k(u_k, f_k, q_k) \\
  \text{Measurement equation: } d_k = h_k(u_k, c_k)
  \]  

  (1)

- Bayesian parameter inference:

  \[
  p(\phi|D, M_i) = \frac{p(\phi|M_i)p(D|\phi, M_i)}{p(D|M_i)} \propto p(\phi|M_i)p(D|\phi, M_i)
  \]  

  (2)

- Bayesian model selection
  - Posterior model probability:

    \[
    P(M_i|D, M) = \frac{p(D|M_i)P(M_i|M)}{p(D|M)}
    \]  

    (3)

  - Evidence-based model selection embodies quantitative Ockham’s razor:

    \[
    \ln p(D|M_i) = \int p(D|\phi)p(\phi)d\phi = \mathbb{E}[\ln p(D|\phi, M_i)] - \mathbb{E} \left[ \ln \frac{p(D, M_i)}{p(\phi|M_i)} \right]
    \]  

    (4)

  - Bayesian vs frequentist [Complexity = function of $p$]:

    \[
    -0.5 \times \text{AIC}_i = \ln p(D|\hat{\phi}, M_i) - p \\
    -0.5 \times \text{BIC}_i = \ln p(D|\hat{\phi}, M_i) - \frac{\ln n}{2}p
    \]  

    (5)
Bayesian inversion methodology

- Likelihood computation: [Khalil et al., JSV, 2013]

\[ p(\phi|D, M_i) = \frac{p(\phi|M_i)p(D|\phi, M_i)}{p(D|M_i)} \propto p(\phi|M_i)p(D|\phi, M_i) \]

\[ \propto p(\phi|M_i) \prod_{k=1}^{n} \int p(d_k|u_k, \phi) p(u_k|d_{1:k-1}, \phi) du_k \]

- Posterior predictive distribution of QOI \( y \):

\[ p(y|D, M) = \sum_{i=1}^{P} \left\{ \int p(y|\phi, D, M_i)p(\phi|D, M_i)d\phi \right\} P(M_i|D, M) - \text{Model averaging} \]

\[ \approx \int p(y|\phi, D, M^*)p(\phi|D, M^*)d\phi - \text{Model selection} \quad \text{[if } P(M^*|D, M) \approx 1] \]

Let's take an example: Nonlinear aeroelastic oscillator
Example: Nonlinear aeroelastic oscillator

- Rigid, unswept, untapered NACA0012 wing installed in the low-speed wind-tunnel at Royal Military College of Canada. Only pitch ($\theta$) degree-of-freedom allowed.

Sandhu et al., JCP, 2016
Example: Nonlinear aeroelastic oscillator

- Self-sustained pitch oscillations were observed in the transitional regime $5.0 \times 10^4 \leq R \leq 1.3 \times 10^5$ wherein the dynamic instability (negative aerodynamic damping) leads to diverging amplitude and the subsequent stabilization (nonlinear aerodynamics) leads to a limit cycle oscillation (LCO).

  ![Graph of self-sustained pitch oscillations](image1)

  Sandhu et al., JCP, 2016

- The nonlinear and unsteady aerodynamics are physically related to the presence of the laminar separation bubble on the wing surface. [Poirel et al., JFS, 2010]

  ![Diagram of nonlinear aerodynamics](image2)

  Sandhu et al., JCP, 2016
Example: Nonlinear aeroelastic oscillator

- Inverse problem for modelling the aeroelastic oscillator: [Sandhu et al., JCP, 2016]
  - Equation of motion:
    \[
    I_{EA} \ddot{\theta} + D \dot{\theta} + K \theta + K' \theta^3 = D' \text{sign}(\dot{\theta}) + \frac{1}{2} \rho U^2 c^2 s C_M(\theta, \dot{\theta}, \ddot{\theta})
    \] (8)
  - Proposed representation of (nonlinear and unsteady) aerodynamic loads \((C_M)\):
    \[
    \mathcal{M}_1 : \quad C_M = e_1 \theta + e_2 \dot{\theta} + e_3 \theta^3 + e_4 \theta^2 \dot{\theta} + \sigma \xi(\tau)
    \]
    \[
    \mathcal{M}_2 : \quad C_M = e_1 \theta + e_2 \dot{\theta} + e_3 \theta^3 + e_4 \theta^2 \dot{\theta} + e_5 \theta^5 + \sigma \xi(\tau)
    \]
    \[
    \mathcal{M}_3 : \quad \frac{\dot{C}_M}{B} + C_M = e_1 \theta + e_2 \dot{\theta} + e_3 \theta^3 + e_4 \theta^2 \dot{\theta} + \frac{c_6}{B} \ddot{\theta} + \sigma \xi(\tau)
    \]
    \[
    \mathcal{M}_4 : \quad \frac{\dot{C}_M}{B} + C_M = e_1 \theta + e_2 \dot{\theta} + e_3 \theta^3 + e_4 \theta^2 \dot{\theta} + e_5 \theta^5 + \frac{c_6}{B} \ddot{\theta} + \sigma \xi(\tau)
    \]
    \[
    \mathcal{M}_5 : \quad \frac{\dot{C}_M}{B_1 B_2} + \frac{(B_1 + B_2) \dot{C}_M}{B_1 B_2} + C_M = e_1 \theta + e_2 \dot{\theta} + e_3 \theta^3 + e_4 \theta^2 \dot{\theta} + \frac{c_1}{B_1 B_2} \ddot{\theta} + \frac{c_2}{B_1 B_2} \dddot{\theta} + \sigma \xi(\tau)
    \]
  - Measurement equation:
    \[
    d_k = \theta_k + \epsilon_k \quad ; \quad k = 1, \ldots, n_d
    \] (9)
Example: Nonlinear aeroelastic oscillator

- Bayesian model selection results:

<table>
<thead>
<tr>
<th>$\mathcal{M} \in \mathcal{M}$</th>
<th>Goodness-of-fit</th>
<th>EIG</th>
<th>Log-evidence</th>
<th>Posterior model probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{M}_1$</td>
<td>35.73</td>
<td>19.35</td>
<td>16.37</td>
<td>0.000</td>
</tr>
<tr>
<td>$\mathcal{M}_2$</td>
<td>36.20</td>
<td>20.50</td>
<td>15.70</td>
<td>0.000</td>
</tr>
<tr>
<td>$\mathcal{M}_3$</td>
<td>132.74</td>
<td>19.09</td>
<td>113.65</td>
<td>0.845</td>
</tr>
<tr>
<td>$\mathcal{M}_4$</td>
<td>132.50</td>
<td>20.55</td>
<td>111.95</td>
<td>0.155</td>
</tr>
<tr>
<td>$\mathcal{M}_5$</td>
<td>123.79</td>
<td>44.67</td>
<td>79.13</td>
<td>0.000</td>
</tr>
<tr>
<td>$\mathcal{M}_6$</td>
<td>123.74</td>
<td>45.64</td>
<td>78.10</td>
<td>0.000</td>
</tr>
<tr>
<td>$\mathcal{M}_7$</td>
<td>123.89</td>
<td>54.59</td>
<td>69.30</td>
<td>0.000</td>
</tr>
<tr>
<td>$\mathcal{M}_8$</td>
<td>121.31</td>
<td>40.84</td>
<td>80.47</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- Practical issues with Bayesian model selection?
  - How to assign parameter prior pdfs to model parameters with limited knowledge (like $e_5$ in $\mathcal{M}_2$)?
  - Are we missing out on better models? Can data point towards a better model?

  Possible solution: SBL/ARD.
Sparse bayesian learning (SBL)
Sparse bayesian learning

- First proposed in Machine learning literature:


- A Bayesian algorithm to obtain sparse feature/parameter space for supervised learning (regression or classification) applications.
- Learning process of support vector machine (SVM) models is posed in Bayesian framework using SBL, leading to relevance vector machine (RVM).

<table>
<thead>
<tr>
<th>Model</th>
<th>Algorithm for obtaining sparse $\mathbf{w}$</th>
</tr>
</thead>
</table>
| SVM   | minimize the model error  
|       | maximize the margin between two classes |
| RVM   | Bayesian inversion methodology  
|       | Sparse Bayesian learning |
Sparse bayesian learning

- Sparse Bayesian learning methodology:

  Bayesian linear regression with conjugate priors

<table>
<thead>
<tr>
<th>Observations</th>
<th>$y = {y_1, y_2, \ldots, y_M}^T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>$y = \Phi w + \epsilon$ [PCE, Radial basis, RVM]</td>
</tr>
<tr>
<td>Obs. Noise</td>
<td>$p(\epsilon) = \mathcal{N}(0, I_M/\rho)$</td>
</tr>
<tr>
<td>ARD Prior for $w$</td>
<td>$p(w</td>
</tr>
<tr>
<td>Prior for $\alpha_i$ and $\rho$</td>
<td>$p(\alpha_i) = \mathcal{G}(\alpha_i</td>
</tr>
</tbody>
</table>

| Likelihood function  | $p(y|w, \rho) = \mathcal{N}(y|\Phi w, I_M/\rho)$ |
| Posterior parameter pdf | $p(w|y, \rho) = \mathcal{N}(w|\rho P^{-1} \Phi^T y, (A + \rho \Phi^T \Phi)^{-1}) = \mathcal{N}(w|m, P)$ |
| Evidence             | $p(y|\alpha, \rho) = \mathcal{N}(y|0, \Phi A^{-1} \Phi^T + I_M/\rho)$ |
| Predictive distribution | $p(y^*|y) = \mathcal{N}(y^*|\Phi^* m, (\Phi^* P(\Phi^*)^T) + I_M/\rho)$ |

- Hierarchical structure of SBL:

  ![Hierarchical structure of SBL](image)
Example: Polynomial regression

- Using Hierarchical Bayes approach, posterior distribution of hyper-parameters is computed as:

\[ p(\alpha, \rho | y) = \frac{p(y | \alpha, \rho) p(\alpha, \rho)}{p(y)} \]  

(10)

- Sparse solution to \( w \) (or basis space \( \Phi \)) pertains to optimal hyperparameter value (Mode of \( p(\alpha, \rho | y) \)) obtained by maximizing model evidence:

\[ (\alpha^*, \rho^*) = \arg \max \{ p(\alpha, \rho | y) \} \]

\[ = \arg \max \{ p(y | \alpha, \rho) p(\alpha) p(\rho) \} \]

- Semi-analytical solution exists!

\[ \alpha_{i}^{new} = \frac{1 - \alpha_{i} P_{i,i} + 2a}{m_{i}^2 + 2b} \]

\[ \rho_{new} = M - \sum (1 - \alpha_{i} P_{i,i}) + 2c \]

\[ \frac{(y - \Phi m)^T (y - \Phi m) + 2d}{(y - \Phi m)^T (y - \Phi m) + 2d} \]

Let’s take an example of polynomial regression
Example: Polynomial regression

- Let say noisy observations $y$ are generated from $y_i = 1.0 + x_i^3 + \epsilon_i$, where $\epsilon \sim \mathcal{N}(0, 1/60)$ is additive gaussian white noise process.

- We propose the following overly complex model to represent the underlying system:

  Model: $y_i = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 + a_4 x_i^4 + a_5 x_i^5 + \epsilon_i$

  ARD Prior: $p(w|\alpha) = \prod_{i=0}^{5} \mathcal{N}(a_i|0, 1/\alpha_i)$ (11)
SBL in action

This is what happens when we have complex models with uninformative priors!
**SBL in action**

### Context of this talk
- Bayesian inversion methodology
- Sparse bayesian learning (SBL)
- SBL for nonlinear dynamical systems
- Application: Nonlinear aeroelastic oscillator
- Conclusion and future direction

### SBL iteration: 2
- $\alpha_0 = 1.02$
- $\alpha_1 = 18.38$
- $\alpha_2 = 263.77$
- $\alpha_3 = 8.77$
- $\alpha_4 = 76.56$
- $\alpha_5 = 3.27$
- $\rho = 51.74$
- Log-evid = 21.10
SBL in action

SBL iteration: 5
\( \alpha_0 = 1.02 \)
\( \alpha_1 = 26.16 \)
\( \alpha_2 = 9228 \)
\( \alpha_3 = 4.69 \)
\( \alpha_4 = 1670 \)
\( \alpha_5 = 5.31 \)
\( \rho = 52.25 \)
Log-evid = 22.68
SBL for nonlinear dynamical systems
SBL for nonlinear dynamical systems

- We want to model the nonlinear and unsteady aerodynamics of the LCO using wind-tunnel data.

- Reformulate the inverse problems as: Given the following encompassing model $\hat{\mathcal{M}}$ of the aerodynamic loading, find the best model nested under this model?.

\[
\frac{\dot{C}_M}{B} + C_M = a_1 \dot{\theta} + a_2 \dot{\theta}^3 + a_4 \dot{\theta}^5 + \sigma \xi(\tau)
\]  

But I want to include my prior knowledge!
SBL for nonlinear dynamical systems

- Model parameters are categorized based on prior knowledge about the aerodynamics as Required \( \phi_{\psi} \) or Contentious \( \phi_{\psi} \)

\[
\frac{\dot{C}_M}{B} + C_M = a_1 \theta + a_2 \dot{\theta} + a_3 \theta^3 + a_4 \theta^2 \dot{\theta} + a_5 \theta^5 + a_6 \theta^4 \dot{\theta} + \frac{c_6}{B} \ddot{\theta} + \sigma \xi(\tau)
\]  

(13)

- ARD prior (from SBL) is only assigned to parameters with limited prior knowledge while other parameters are assigned known priors.

| Prior pdf, \( p(\phi | \psi) = p(\phi_{\psi}) \); Hyper-parameter, \( \psi = \{ \alpha_1, \alpha_2, \alpha_3, \alpha_4 \} \) |
|----------------------------------|
| \( p(\phi_{\psi}) \) | \( \mathcal{L}(B|0.2, 50) \) \( U(a_1|-2, 0) \) \( U(a_2|-2, 0) \) \( \mathcal{L}(\sigma|0.002, 50) \) |
| \( p(\phi_{\psi} | \psi) \) | ARD prior, \( \mathcal{N}(a_3|0, \frac{1}{\alpha_1}) \) \( \mathcal{N}(a_4|0, \frac{1}{\alpha_2}) \) \( \mathcal{N}(a_5|0, \frac{1}{\alpha_3}) \) \( \mathcal{N}(a_6|0, \frac{1}{\alpha_4}) \) |

Sandhu et al., CMAME, 2017

- Discrete model space is converted into a continuous model space with the use of ARD prior.
SBL for nonlinear dynamical systems

- Hierarchical structure of Bayesian inference/SBL:

  ![Hierarchical Diagram]

  - Hyper-parameter \( \psi \rightarrow \) Prior pdf → Model parameter \( \phi \)
  - Observations \( D \) → Measurement Eqn → State vector \( u \)

- Implementation of SBL for nonlinear dynamical systems:
  - Identify sparse model parameter space through evidence maximization: Task: Optimization
    \[
    \psi_{\text{map}} = \arg \max \{ p(D|\psi) \} \tag{14}
    \]
  - Model evidence as a function of hyper-parameter Task: Evidence computation
    \[
    p(D|\psi) = \int p(D|\phi)p(\phi|\psi)d\phi \tag{15}
    \]
  - Likelihood computation Task: State estimation
    \[
    p(D|\phi) = \prod_{k=1}^{M} \int p(d_k|u_{j(k)}, \phi)p(u_{j(k)}|d_{1:k-1}, \phi)du_{j(k)} \tag{16}
    \]
SBL for nonlinear dynamical systems

- Numerical implementation:
  - Evidence optimization: Derivative-free methods including line-search, pattern search, simplex method, evolutionary algorithms; and many others.
  - Evidence computation: Chib-Jeliazkov method, Transitional MCMC, Power posteriors, Nested sampling, Annealed importance sampling, Harmonic mean estimator, Gauss-Hermite quadrature; and many others.
  - MCMC sampler for Chib-Jeliazkov method: Metropolis-Hastings, Gibbs, TMCMC, adaptive Metropolis, Delayed Rejection Adaptive Metropolis (DRAM); and many others.
  - State estimation: Kalman filter, Extended Kalman filter, unscented Kalman filter, ensemble Kalman filter, particle filter; and many others.
Application: Nonlinear aeroelastic oscillator
Numerical results using synthetic data

- We generate a synthetic LCO using the following model:

\[ \frac{\dot{C}_M}{B} + C_M = a_1 \theta + a_2 \dot{\theta} + a_3 \theta^3 + a_4 \theta^2 \dot{\theta} + \frac{c_6}{B} \ddot{\theta} + \sigma \xi(\tau) \]  

(17)

- Unidimensional ARD prior for relevant parameter
  - Model proposed (= data-generating model) and assuming we don’t know anything about \( a_3 \theta^3 \) term:

\[ \frac{\dot{C}_M}{B} + C_M = a_1 \theta + a_2 \dot{\theta} + a_3 \theta^3 + a_4 \theta^2 \dot{\theta} + \frac{c_6}{B} \ddot{\theta} + \sigma \xi(\tau) \]  

(18)

- Two choices of ARD priors are considered (Gauss vs Laplace ARD priors):
  1. \( p(\phi|\psi) = L(B|0.2, 50) U(a_1|-2, 0) U(a_2|-2, 0) N(a_3|0, 1/\alpha) U(a_4|-600, 0) L(\sigma|0.002, 50) \)
  2. \( p(\phi|\psi) = L(B|0.2, 50) U(a_1|-2, 0) U(a_2|-2, 0) LP(a_3|0, 1/\alpha) U(a_4|-600, 0) L(\sigma|0.002, 50) \)
Numerical results using synthetic data

- Hyper-parameter ($\alpha$) vs Bayesian entities:
  - Change in log-evidence driven by loss of goodness-of-fit due to removal of $a_3$
  - Log-evidence has higher slope near maxima and is minimally sloped elsewhere
  - Both Laplace prior and Gaussian prior results in same parameter sparsity level.

Sandhu et al., CMAME, 2017
Numerical results using synthetic data

- Effect of using zero-mean ARD priors on posterior parameter estimates:

Sandhu et al., CMAME, 2017
Numerical results using synthetic data

- Unidimensional ARD prior for irrelevant parameter
  - Model proposed (additional term $a_5 \theta^5$ than the data-generating model):
    \[
    \frac{\dot{C}_M}{B} + C_M = a_1 \dot{\theta} + a_2 \dot{\theta} + a_3 \theta^3 + a_4 \theta^2 \dot{\theta} + a_5 \theta^5 + \frac{c_6}{B} \ddot{\theta} + \sigma \xi(\tau)
    \]  
    (19)

  - Prior pdf:
    \[
    p(\phi|\psi) = \mathcal{L}(B|0.2, 50) \mathcal{U}(a_1|-2, 0) \mathcal{U}(a_2|-2, 0) \mathcal{U}(a_3|-250, 250) \mathcal{U}(a_4|-600, 0) \mathcal{N}(a_5|0, 1/\alpha) \mathcal{L}(\sigma|0.002, 50)
    \]

  - Observations:
    - Increase in log-evidence driven by decrease in complexity (removal of $a_5$)
    - Log-evidence is flat in regions higher the optimal hyperparameter (Issues with optimization!)

Sandhu et al., CMAME, 2017
Numerical results using synthetic data

- Multiple parameters (mix of relevant and irrelevant parameters) with ARD priors:
  - Model proposed:
    \[
    \frac{\ddot{C}_M}{B} + C_M = a_1 \theta + a_2 \dot{\theta} + a_3 \theta^3 + a_4 \theta^2 \dot{\theta} + a_5 \theta^5 + a_6 \theta^4 \dot{\theta} + \frac{c_6}{B} \ddot{\theta} + \sigma \xi(\tau)
    \]  
  - Prior pdf:
    - Prior pdf, \( p(\phi|\psi) = p(\phi_{\psi})p(\phi|\psi) \); Hyper-parameter, \( \psi = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} \)
    - \( p(\phi_{\psi}) \):
      \[
      \mathcal{L}(B|0.2, 50) \mathcal{U}(a_1|-2, 0) \mathcal{U}(a_2|-2, 0) \mathcal{L}(\sigma|0.002, 50)
      \]
    - \( p(\phi|\psi) \):
      - ARD prior, \( \mathcal{N}(a_3|0, \frac{1}{\alpha_1}) \mathcal{N}(a_4|0, \frac{1}{\alpha_2}) \mathcal{N}(a_5|0, \frac{1}{\alpha_3}) \mathcal{N}(a_6|0, \frac{1}{\alpha_4}) \)

Sandhu et al., CMAME, 2017
Numerical results using synthetic data

- SBL results:

Sandhu et al., CMAME, 2017
Numerical results using synthetic data

- Comparison of marginal and joint posterior pdf of relevant parameters $a_3$ and $a_4$ for ARD prior with optimal hyper-parameters and flat prior pdf:

Sandhu et al., CMAME, 2017
Conclusion and future direction
Conclusion

- **Conclusion:**
  - Machine learning algorithm of SBL/ARD is exploited as an automatic model reduction tool within the Bayesian inversion framework, with an application to nonlinear dynamical systems.
  - SBL/ARD approach is validated using a synthetically generated nonlinear aeroelastic oscillations.
  - Both Laplace and Gaussian ARD prior produced same parameter sparsity levels.

- **Future direction:**
  - Apply the SBL/ARD model reduction tool to model wind-tunnel data
  - Using gradient/hessian information to expedite the optimization of model evidence.

- **Full details:**
  

- **SBL polynomial regression code:**

  https://github.com/rimplesandhu/UQMLcodes.git
Acknowledgement

- Financial support:
  - Department of National Defence, Canada
  - Ontario Graduate Scholarship program
  - Carleton University, Ottawa, Canada
  - Natural Sciences and Engineering Research Council, Canada

- Computing resources:
  - HP Linux Cluster, Carleton University, Ottawa, Canada
  - CLUMEQ, McGill University, Montreal, Canada
  - SciNet (Now Niagara), University Of Toronto, Toronto, Canada

- External libraries used:
  - Dakota, UQ Toolkit (UQTk) [by Sandia National Lab]
  - Armadillo (Linear algebra library for C++)

Thank you