This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Extending relevance vector machine to resolve overfitting during Bayesian calibration in nonlinear stochastic dynamics

Rimple Sandhu[1, 2], Mohammad Khalil[3], Chris Pettit[4], Dominique Poirel[5], Abhijit Sarkar[1]

[1]Carleton University, Ottawa, ON, Canada

[2]Currently at National Renewable Energy Laboratory, Golden, CO, USA

[3]Sandia National Laboratories, Livermore, CA, USA

[4]United States Naval Academy, Annapolis, MD, USA

[5]Royal Military College of Canada, Kingston, ON, Canada

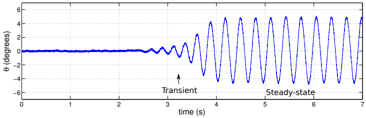
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

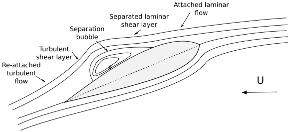
September 17, 2021

boratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, Ll Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE

Nonlinear aeroelastic oscillator - overfitting issue

- Aeroelastic limit-cycle oscillation (LCO) (Sandhu et. al, CMAME, 2014; Sandhu et. al, JCP, 2016)





Sandhu et. al, CMAME, 2014; Sandhu et. al, JCP, 2016 (with copyright permission)

- Identify the best model for C_M:

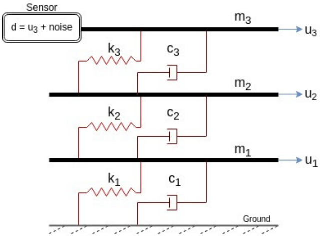
$$\mathcal{M}_1: \quad C_M = e_1\theta + e_2\dot{\theta} + e_3\theta^3 + e_4\theta^2\dot{\theta} + \sigma\xi(\tau)$$

$$\mathcal{M}_2: \quad \frac{\dot{C}_M}{B} + C_M = e_1\theta + e_2\dot{\theta} + e_3\theta^3 + e_4\theta^2\dot{\theta} + \frac{c_6}{B}\ddot{\theta} + \sigma\xi(\tau)$$

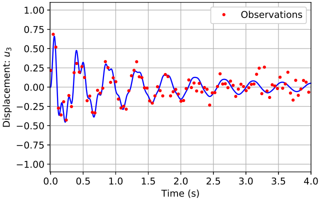
$$\mathcal{M}_3: \quad \frac{\ddot{C}_M}{B_1B_2} + \frac{(B_1+B_2)\dot{C}_M}{B_1B_2} + C_M = e_1\theta + e_2\dot{\theta} + e_3\theta^3 + e_4\theta^2\dot{\theta} + \frac{(2c_6c_7+0.5)\ddot{\theta}}{B_1B_2} + \frac{c_6\dddot{\theta}}{B_1B_2} + \sigma\xi(\tau)$$

Problem definition: Mass-spring-damper system

- Inverse problem: Estimate a probability density function (pdf) of damping and stiffness parameters while identifying the sparsity in damping parameters. (Sandhu *et al.*, JCP, 2021)



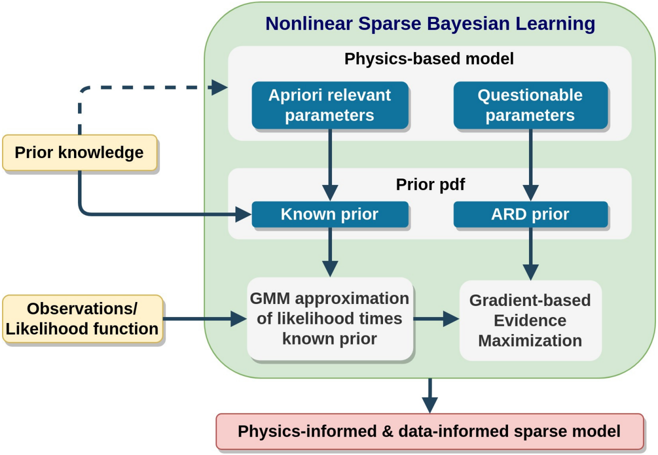
(a) Three-dof mass-spring-damper model.



(b) Measured versus true response.

Sandhu et. al, JCP, 2021 (with copyright permission)

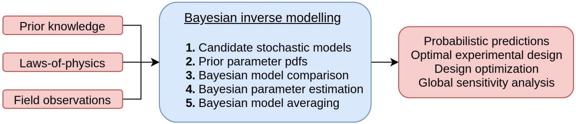
Nonlinear sparse Bayesian learning



Sandhu et. al, JCP, 2021 (with copyright permission)

Bayesian inverse modelling

- Overview of the Bayesian framework



- **Bayesian model comparison**: Ranking of plausible models is based on model evidence,

$$p(\mathcal{D}) = \int p(\mathcal{D}|\phi)p(\phi)d\phi \quad , \quad \underbrace{\ln p(\mathcal{D})}_{\text{Log-evidence}} = \underbrace{E[\ln p(\mathcal{D}|\phi)]}_{\text{Data-fit}} - \underbrace{E\left[\ln \frac{p(\phi|\mathcal{D})}{p(\phi)}\right]}_{\text{Model complexity}}$$

Nonlinear sparse Bayesian learning (NSBL): Methodology

- Stochastic state-space model structure:

$$\text{Model equation}: \quad \mathbf{u}_{k+1} = \mathbf{g}_k(\mathbf{u}_k, \mathbf{f}_k, \mathbf{q}_k; \phi)$$
$$\text{Measurement equation}: \quad \mathbf{d}_k = \mathbf{h}_k(\mathbf{u}_k, \epsilon_k; \phi)$$

- Parameter prior pdf: hybrid assignment ($\boldsymbol{\alpha}$: Hyper-parameters)

$$\phi = \{\phi_\alpha, \phi_{\text{-}\alpha}\} \;\; ; \;\; p(\phi|\boldsymbol{\alpha}) = p(\phi_{\text{-}\alpha})p(\phi_\alpha|\boldsymbol{\alpha}) = \underbrace{p(\phi_{\text{-}\alpha})}_{\text{Known prior}} \underbrace{\prod \mathcal{N}\left(\phi_i|0, \alpha_i^{-1}\right)}_{\text{ARD prior}}$$

- Entity $p(\mathcal{D}|\phi)\, p(\phi_{\text{-}\alpha})$ remains unchanged during sparse learning!

$$p(\phi|\mathcal{D}, \boldsymbol{\alpha}) = \frac{p(\mathcal{D}|\phi)p(\phi|\boldsymbol{\alpha})}{p(\mathcal{D}|\boldsymbol{\alpha})} \propto \underbrace{p(\mathcal{D}|\phi)\, p(\phi_{\text{-}\alpha})}_{\text{Independent of } \boldsymbol{\alpha}} \mathcal{N}(\phi_\alpha|\mathbf{0}, \mathbf{A}^{-1})$$

- Gaussian mixture model (GMM) approximation:

$$p(\mathcal{D}|\phi)p(\phi_{\text{-}\alpha}) \approx \sum_{k=1}^{K} a^{(k)} \mathcal{N}(\phi|\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}),$$

NSBL: Semi-analytical Bayesian framework

Entity	Solution
Parameter decomposition	$\phi = \{\phi_{\alpha}, \phi_{\text{-}\alpha}\}$
Likelihood + known prior	$p(\mathcal{D}\,\vert\,\phi)p(\phi_{\text{-}\alpha}) \approx \sum_{k=1}^{K} a^{(k)} \mathcal{N}(\phi\,\vert\,\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)})$
ARD prior	$p(\phi_{\alpha}\,\vert\,\boldsymbol{\alpha}) = \mathcal{N}(\phi_{\alpha}\,\vert\,\mathbf{0}, \mathbf{A}^{-1})$
Hyper-parameter prior	$p(\boldsymbol{\alpha}) = \prod_{i=1}^{N_{\alpha}} \mathcal{G}(\alpha_i\,\vert\,r_i, s_i)$
Model evidence	$\hat{p}(\mathcal{D}\,\vert\,\boldsymbol{\alpha}) = \sum_{k=1}^{K} a^{(k)} \mathcal{N}(\boldsymbol{\mu}_{\alpha}^{(k)}\,\vert\,\mathbf{0}, \mathbf{B}_{\alpha}^{(k)})$; $\mathbf{B}_{\alpha}^{(k)} = \boldsymbol{\Sigma}^{(k)} + \mathbf{A}^{-1}$
Parameter posterior PDF	$\hat{p}(\phi\,\vert\,\mathcal{D}, \boldsymbol{\alpha}) = \sum_{k=1}^{K} w^{(k)} \mathcal{N}(\phi\,\vert\,\mathbf{m}^{(k)}, \mathbf{P}^{(k)})$; $w^{(k)} = \dfrac{a^{(k)} \mathcal{N}(\boldsymbol{\mu}_{\alpha}^{(k)}\,\vert\,\mathbf{0}, \mathbf{B}_{\alpha}^{(k)})}{\hat{p}(\mathcal{D}\,\vert\,\log \boldsymbol{\alpha})}$
Objective function	$\mathcal{L}(\log \boldsymbol{\alpha}) = \log \hat{p}(\mathcal{D}\,\vert\,\log \boldsymbol{\alpha}) + \sum_{i=1}^{N_{\alpha}} (r_i \log \alpha_i - s_i \alpha_i)$
Gradient of $\mathcal{L}(\log \boldsymbol{\alpha})$	$g_i(\log \boldsymbol{\alpha}) = \sum_{k=1}^{K} w^{(k)} v_i^{(k)} + r_i - s_i \alpha_i = \bar{v}_i + r_i - s_i \alpha_i$ $v_i^{(k)} = (\gamma_i^{(k)} - \alpha_i (m_i^{(k)})^2)/2$; $\gamma_i^{(k)} = 1 - \alpha_i P_{ii}^{(k)}$
Hessian of $\mathcal{L}(\log \boldsymbol{\alpha})$	$H_{ij}(\log \boldsymbol{\alpha}) = \sum_{k=1}^{K} w^{(k)} \left[\alpha_i \alpha_j \left(\dfrac{(P_{ij}^{(k)})^2}{2} + m_i^{(k)} m_j^{(k)} P_{ij}^{(k)} \right) + v_i^{(k)} v_j^{(k)} - \bar{v}_i \bar{v}_j \right] + \delta_{ij} \left[\bar{v}_i - \dfrac{1}{2} - s_i \alpha_i \right]$

Sandhu et. al, JCP, 2021 (with copyright permission)

NSBL: Numerical Implementation

- Stage 1: GMM construction by sampling $p(\mathcal{D}|\phi)p(\phi_{\text{-}\alpha})$ and then utilizing
  - Kernel density estimation (KDE): Computationally expedient, # of kernels = # of samples (Our choice)
  - Expectation maximization (EM): More involved, smaller # of kernels (Not pursued here)
- Stage 2: Newton's method to maximize model evidence (**non-convex optimization**):

$$\log \boldsymbol{\alpha}_{j+1} = \log \boldsymbol{\alpha}_j + \beta_j \mathbf{p}_j \ , \text{ where } \ \mathbf{H}_j \mathbf{p}_j = -\mathbf{g}_j$$

  - Modified Newton method
  - Trust-region Newton method (Our choice)
- Stage 3: Sparsity identification by exploiting relevance indicator, defined as

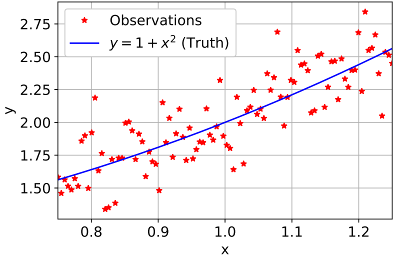
$$\gamma_i^{(k)} = 1 - \frac{\alpha_i}{(P_{ii}^{(k)})^{-1}} = 1 - \frac{\alpha_i}{\alpha_i + \{(\boldsymbol{\Sigma}_\alpha^{(k)})^{-1}\}_{(i,i)}} \in [0, 1], \qquad (1)$$

NSBL: Benefits

- Benefits of NSBL over standard Bayesian model updating
  - Evidence-based pruning of redundant parameters while incorporating prior parametric knowledge
  - Parameter posterior pdf and model evidence available semi-analytically
  - Gradient and Hessian of evidence with respect to hyper-parameters available analytically
  - Can handle non-Gaussian and multimodal pdfs

Example: Polynomial Regression with Multimodal Prior

- Generated observations using $y_i = 1 + x_i^2 + \epsilon_i$, where ϵ_i is a Gaussian white noise process with distribution $\mathcal{N}(\epsilon_i|0, 0.04)$. (Sandhu *et al.*, JCP, 2021)



Sandhu et. al, JCP, 2021 (with copyright permission)

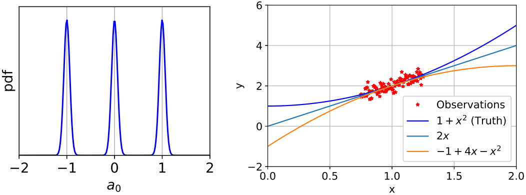
- The observational data consists of 100 noisy samples with equally spaced coordinate x within the domain $x \in [0.75, 1.25]$

Example: Polynomial Regression with Multimodal Prior

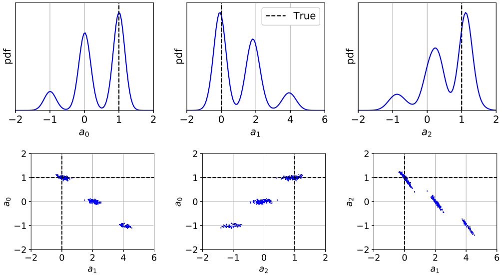
- Inverse problem setup: (Sandhu *et al.*, JCP, 2021, with copyright permission)

Proposed model	$y = a_0 + a_1 x + a_2 x^2 + \epsilon \;\; ; \;\; \epsilon \sim \mathcal{N}(0, 0.04)$
ϕ decomposition	$\phi_\alpha = \{a_1, a_2\} \; , \quad \phi_{\text{-}\alpha} = \{a_0\}$
ARD prior, $p(\phi_\alpha \vert \alpha)$	$\mathcal{N}(a_1 \vert 0, \alpha_1^{-1}) \; \mathcal{N}(a_2 \vert 0, \alpha_2^{-1})$
Known prior, $p(\phi_{\text{-}\alpha})$	$\left[\mathcal{N}(a_0 \vert \text{-}1, 0.2^2) + \mathcal{N}(a_0 \vert 0, 0.2^2) + \mathcal{N}(a_0 \vert 1, 0.2^2)\right] / 3$

- Non-Gaussian prior knowledge: (Sandhu *et al.*, JCP, 2021, with copyright permission)



- Stage 1 (GMM construction): 2500 samples distributed according to $p(\mathcal{D}|\phi)p(\phi_{-\alpha})$ or $p(\mathcal{D}|a_0, a_1, a_2)p(a_0)$ using TMCMC, followed by a GMM construction using KDE (Sandhu *et al.*, JCP, 2021, with copyright permission)

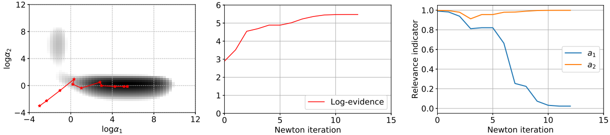


SBL and BCS are not applicable for multimodal/Non-Gaussian posterior pdfs!

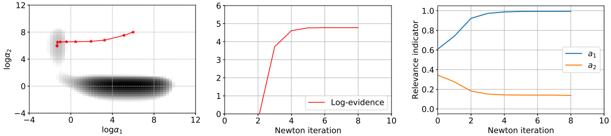
Example: Polynomial Regression with Multimodal Prior

Stage 2: Sparsification (Sandhu *et al.*, JCP, 2021, with copyright permission)

- Initialized at $\log \boldsymbol{\alpha} = \{-3, -3\}$, NSBL suggests a_1 is irrelevant and a_2 is relevant.



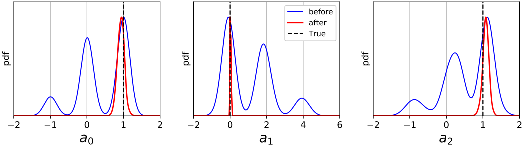
- Iteration initialized at $\log \boldsymbol{\alpha} = \{6, 8\}$, NSBL suggests a_1 is relevant and a_2 is irrelevant.



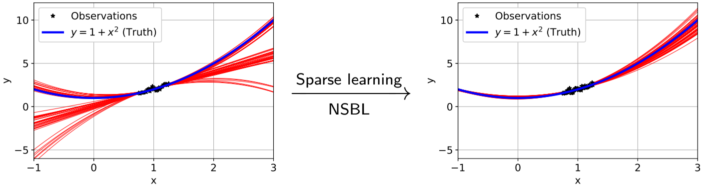
Multistart is necessary!

Example: Polynomial Regression with Multimodal Prior

- Posterior parameter pdf post sparse learning: (Sandhu *et al.*, JCP, 2021, with copyright permission)

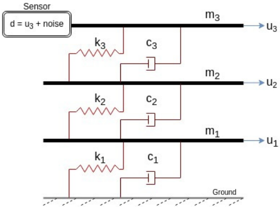


- Predictions before and after sparse learning: (Sandhu *et al.*, JCP, 2021, with copyright permission)

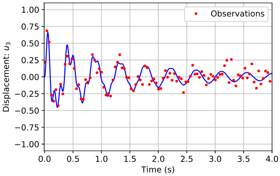


Example: Mass-spring-damper system

- Displacement vector $\mathbf{u} = \{u_1, u_2, u_3\}$; equation-of-motion $\mathbf{M\ddot{u}} + \mathbf{C\dot{u}} + \mathbf{Ku} = 0$ (Sandhu *et al.*, JCP, 2021, with copyright permission)



(c) Three-dof mass-spring-damper model.



(d) Measured versus true response.

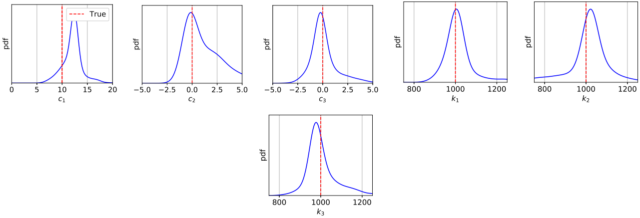
- Parameter true values:
  - Mass: $m_1 = m_2 = m_3 = 1.0$,
  - Stiffness: $k_1 = k_2 = k_3 = 1000.0$,
  - Damping: $c_1 = 10$, $c_2 = 0$, $c_3 = 0$.

- Inverse problem setup: (Sandhu *et al.*, JCP, 2021, with copyright permission)

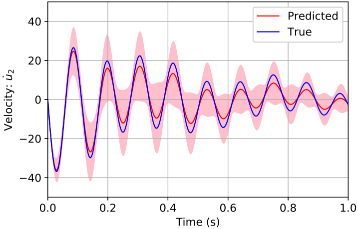
Proposed model	$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{0}$
ϕ decomposition	$\phi_\alpha = \{c_1, c_2, c_3\}$, $\phi_{\text{-}\alpha} = \{k_1, k_2, k_3\}$
Known prior, $p(\phi_{\text{-}\alpha})$	$\mathcal{U}(k_1\|0, 5000)\, \mathcal{U}(k_2\|0, 5000)\, \mathcal{U}(k_3\|0, 5000)$
ARD prior, $p(\phi_\alpha\|\boldsymbol{\alpha})$	$\mathcal{N}(c_1\|0, \alpha_1^{-1})\, \mathcal{N}(c_2\|0, \alpha_2^{-1})\, \mathcal{N}(c_3\|0, \alpha_3^{-1})$
Hyperprior, $p(\boldsymbol{\alpha})$	$\prod_{i=1}^{3} \mathcal{G}(\alpha_i\|r, s)$; $\log r = \log s =$ -10

Example: Mass-spring-damper system

- Bayesian inference with flat priors for questionable parameters: (Sandhu *et al.*, JCP, 2021, with copyright permission)

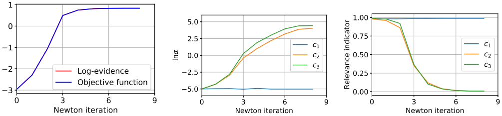


- Predicted versus true time-history of velocity $\dot{u}_2$: (Sandhu *et al.*, JCP, 2021, with copyright permission)

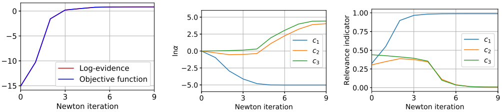


Example: Mass-spring-damper system

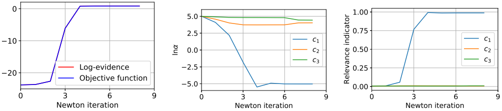
- Stage 2: Sparsification (Sandhu *et al.*, JCP, 2021, with copyright permission)



(e) $\boldsymbol{\alpha}$ initiated at $\{-5, -5, -5\}$



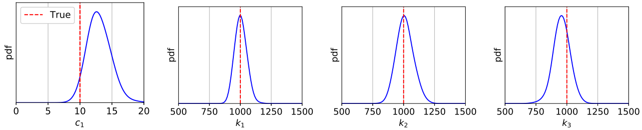
(f) $\boldsymbol{\alpha}$ initiated at $\{0, 0, 0\}$



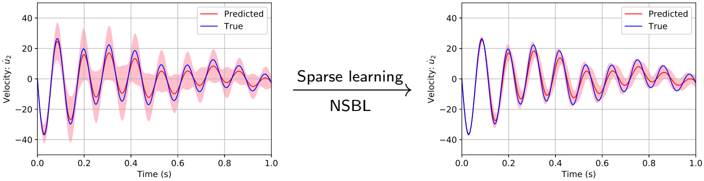
(g) $\boldsymbol{\alpha}$ initiated at $\{5, 5, 5\}$

Example: Mass-spring-damper system

- Bayesian inference post sparse learning, using optimal ARD priors from NSBL: (Sandhu *et al.*, JCP, 2021, with copyright permission)



- Prediction before and after sparse learning of model parameters: (Sandhu *et al.*, JCP, 2021, with copyright permission)



Likelihood and Parameter Posterior

$$\mathbf{x}_{k+1} = \mathbf{g}_k\left(\mathbf{x}_k, \phi, \mathbf{f}_k, \mathbf{q}_k\right) \tag{2}$$

$$\mathbf{d}_j = \mathbf{h}_j(\mathbf{x}_{d(j)}, \varepsilon_j) \tag{3}$$

$$p(\phi|\mathcal{D}) \propto p(\phi) \underbrace{\prod_{j=1}^{J} \int p(\mathbf{d}_j|\mathbf{x}_{d(j)}, \phi) \underbrace{p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \phi)}_{\text{Non-Gaussian filtering}} d\mathbf{x}_{d(j)}}_{p(\mathcal{D}|\phi)}$$

$$p(\phi|\mathcal{D}) \propto p(\phi)p(\mathcal{D}|\phi)$$

Bisaillon et al., Nonlinear Dynamics, 2015

Likelihood Computation using EKF

$$p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\phi) \sim \mathcal{N}(\mathbf{x}_{d(j)}|\mathbf{x}^f_{d(j)},\mathbf{P}^f_{d(j)})$$

$$p(\mathcal{D}|\phi) = \prod_{j=1}^{J} \mathcal{N}(\mathbf{d}_j|\mathbf{h}_j(\mathbf{x}^f_{d(j)},\mathbf{0}),\mathbf{\Sigma}')$$

$$\mathbf{\Sigma}' = \mathbf{C}_j\mathbf{P}^f_{d(j)}\mathbf{C}_j^T + \mathbf{D}_j\mathbf{\Gamma}_j\mathbf{D}_j^T$$

$$\mathbf{C}_j = \left.\frac{\partial \mathbf{h}_j(\mathbf{x}_{d(j)},\varepsilon_j)}{\partial \mathbf{x}_{d(j)}}\right|_{\mathbf{x}_{d(j)}=\mathbf{x}^f_{d(j)},\varepsilon_j=\mathbf{0}} \tag{4}$$

$$\mathbf{D}_j = \left.\frac{\partial \mathbf{h}_j(\mathbf{x}_{d(j)},\varepsilon_j)}{\partial \varepsilon_j}\right|_{\mathbf{x}_{d(j)}=\mathbf{x}^f_{d(j)},\varepsilon_j=\mathbf{0}} \tag{5}$$

Bisaillon et al., Nonlinear Dynamics, 2015

Likelihood Computation using EnKF

$$p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\phi) \approx \frac{1}{N}\sum_{s=1}^{N}\delta(\mathbf{x}_{d(j)} - \mathbf{x}^{f}_{d(j),s}).$$

$$\begin{aligned}
p(\mathcal{D}|\phi) &= \prod_{j=1}^{J}\int_{-\infty}^{\infty} p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\phi)p(\mathbf{d}_j|\mathbf{x}_{d(j)},\phi)d\mathbf{x}_{d(j)} \\
&\approx \prod_{j=1}^{J}\int_{-\infty}^{\infty} \frac{1}{N}\sum_{s=1}^{N}\delta(\mathbf{x}_{d(j)} - \mathbf{x}^{f}_{d(j),s})p(\mathbf{d}_j|\mathbf{x}_{d(j)},\phi)d\mathbf{x}_{d(j)} \\
&\approx \prod_{j=1}^{J}\left[\frac{1}{N}\sum_{s=1}^{N}p(\mathbf{d}_j|\mathbf{x}^{f}_{d(j),s},\phi)\right]
\end{aligned}$$

Bisaillon et al., Nonlinear Dynamics, 2015

Likelihood Computation using PF

$$p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\phi) \approx \sum_{s=1}^{N} w_{d(j),s}\,\delta(\mathbf{x}_{d(j)} - \mathbf{x}^f_{d(j),s})$$

$$\begin{aligned}
p(\mathcal{D}|\phi) &= \prod_{j=1}^{J} \int_{-\infty}^{\infty} p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\phi)\,p(\mathbf{d}_j|\mathbf{x}_{d(j)},\phi)\,d\mathbf{x}_{d(j)} \\
&\approx \prod_{j=1}^{J} \int_{-\infty}^{\infty} \sum_{s=1}^{N} w_{d(j),s}\,\delta(\mathbf{x}_{d(j)} - \mathbf{x}^f_{d(j),s})\,p(\mathbf{d}_j|\mathbf{x}_{d(j)},\phi)\,d\mathbf{x}_{d(j)} \\
&\approx \prod_{j=1}^{J} \sum_{s=1}^{N} w_{d(j),s}\,p(\mathbf{d}_j|\mathbf{x}^f_{d(j),s},\phi)
\end{aligned} \tag{6}$$

$$w_{d(j),i} \propto w_{d(j)-1,i}\,\frac{p(\mathbf{d}_j|\mathbf{x}_{d(j),i})\,p(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i})}{q(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i},\mathbf{d}_j)}$$

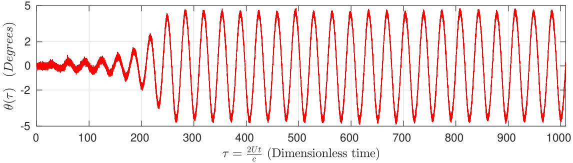
Bisaillon et al., Nonlinear Dynamics, 2015

Single degree-of-freedom aeroelastic system

Data generating model:

$$I\ddot{\theta} + D\dot{\theta} + K\theta + K_{nl}\theta^3 = D_{nl}\,\mathrm{sgn}(\dot{\theta}) + \frac{1}{2}\rho U^2 c^2 s\, C_M(\theta, \dot{\theta}, \ddot{\theta})$$

$$\frac{\dot{C}_M}{B} + C_M = e_1\theta + e_2\dot{\theta} + e_3\theta^3 + e_4\theta^2\dot{\theta} + \frac{c_6}{B}\ddot{\theta} + \sigma\xi(\tau)$$



Sandhu et. al, CMAME 2017 (with copyright permission)

Case 1: One-dimensional sparse learning problem

1. Aerodynamic parameter e_3 is treated as questionable

$$\frac{\dot{C}_M}{B} + C_M = e_1\theta + e_2\dot{\theta} + e_3\theta^3 + e_4\theta^2\dot{\theta} + \frac{c_6}{B}\ddot{\theta} + \sigma\xi(\tau)$$

   i. ARD prior, $\mathrm{p}(\phi_\alpha|\alpha)$

$$\mathrm{p}(\phi_\alpha|\alpha) = \mathcal{N}(e_3|0, \alpha^{-1})$$

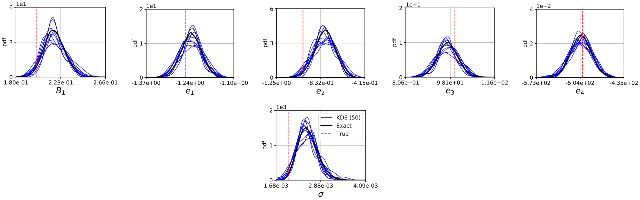
   ii. Known prior, $\mathrm{p}(\phi_{-\alpha})$

$$\mathrm{p}(\phi_{-\alpha}) = \mathcal{L}(B|0.2, 50)\mathcal{U}(e_1|-2, 0)\mathcal{U}(e_2|-2, 0)\mathcal{U}(e_4|-600, 0)\mathcal{L}(\sigma|0.002, 50)$$

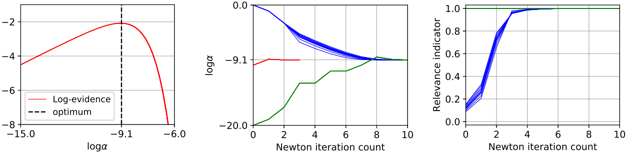
R. Sandhu. Model Comparison and Sparse Learning of Nonlinear Physics-Based Models Using Bayesian Inference. *PhD Thesis*, Carleton University, 2020.

Single degree-of-freedom aeroelastic system

1. Constructing GMM of partial posterior $p(\mathcal{D}|\phi)p(\phi_{-\alpha})$ using KDE



2. Maximizing log evidence with respect to hyperparameter α



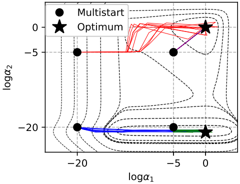
3. Parameter e_3 is correctly identified as relevant

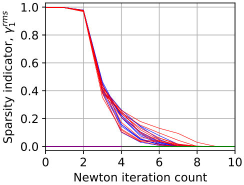
R. Sandhu. Model Comparison and Sparse Learning of Nonlinear Physics-Based Models Using Bayesian Inference. *PhD Thesis*, Carleton University, 2020.

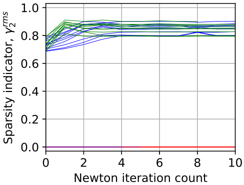
Case 2: Two-dimensional sparse learning problem

1. Aerodynamic parameters e_5 and e_6 are treated as questionable

$$\frac{\dot{C}_M}{B} + C_M = e_1\theta + e_2\dot{\theta} + e_3\theta^3 + e_4\theta^2\dot{\theta} + e_5\theta^5 + e_6\theta^4\dot{\theta} + \frac{c_6}{B}\ddot{\theta} + \sigma\xi(\tau)$$







- (i) One optimum correctly identifies parameters e_5 and e_6 as irrelevant
- (ii) The other finds e_5 to be irrelevant, but e_6 to be relevant

[1] R. Sandhu. Model Comparison and Sparse Learning of Nonlinear Physics-Based Models Using Bayesian Inference. *PhD Thesis*, Carleton University, 2020.

Conclusion and Future Research

- NSBL is effective for ODE models.
- NSBL for coupled nonlinear ODE and PDE models (bending-bending-torsion-pitch motions) using wind-tunnel data.
- Concurrent selection of physics-based model and model errors.
- Address overfitting using NSBL for Bayesian neural networks.
- Apply to fields outside of aeroelasticity (e.g. epidemiology).

Conclusion

Method	Capabilities			Methodology			Scalability
	Nonlinear models?	Prior knowledge?	Sparse learning?	Prior pdf type	Evidence estimation	Evidence optimization	# of likelihood evaluations
SBL/BCS	×	×	✓	Gaussian ARD (Conjugate prior)	Analytical	Gradient set to zero	N/A
Bayesian model comparison	✓	✓	×	Known pdf	Numerical (MCMC-based)	N/A	$N_{\text{Models}} \times N_{\text{MCMC}}$
ARD (Sampling-based)	✓	✓	✓	Known + Gaussian or Laplace ARD	Numerical (MCMC-based)	Gradient-free	$N_{\text{Iter}} \times N_{\text{MCMC}}$
ARD (NSBL)	✓	✓	✓	Known + Gaussian ARD	Semianalytical (GMM-based)	Gradient and Hessian based	N_{MCMC}

N_{MCMC}: # of MCMC samples needed for computing evidence (for an average model).

N_{Models}: # of candidate models.

N_{Iter}: # of iterations required for optimizing evidence.

Sandhu *et al.*, JCP, 2021, with copyright permission

Acknowledgements

Financial contributions

1. Department of Defence, Canada
2. NSERC

Derivation: Parameter Posterior (closely follows the book by Evensen)

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_k, \phi | \mathbf{d}_1, \ldots, \mathbf{d}_J) = \underbrace{p(\mathbf{x}_1, \ldots, \mathbf{x}_k | \mathbf{d}_1, \ldots, \mathbf{d}_J, \phi)}_{\text{Conditional state pdf}} \; \underbrace{p(\phi | \mathbf{d}_1, \ldots, \mathbf{d}_J)}_{\text{Parameter posterior pdf}}$$

$$\begin{aligned}
& p(\mathbf{x}_1, \ldots, \mathbf{x}_k | \mathbf{d}_1, \ldots, \mathbf{d}_J, \phi) p(\phi | \mathbf{d}_1, \ldots, \mathbf{d}_J) \propto \\
& \quad p(\phi) \left[\prod_{j'=1}^{d(1)-1} p(\mathbf{x}_{j'} | \mathbf{x}_{j'-1}, \phi) \right] p(\mathbf{x}_{d(1)} | \mathbf{x}_{d(1)-1}, \phi) p(\mathbf{d}_1 | \mathbf{x}_{d(1)}, \phi) \\
& \qquad \vdots \\
& \quad \left[\prod_{j'=d(J-1)+1}^{d(J)-1} p(\mathbf{x}_{j'} | \mathbf{x}_{j'-1}, \phi) \right] p(\mathbf{x}_{d(J)} | \mathbf{x}_{d(J)-1}, \phi) p(\mathbf{d}_J | \mathbf{x}_{d(J)}, \phi) \\
& \qquad \vdots \\
& \quad \left[\prod_{j'=d(J)+1}^{k} p(\mathbf{x}_{j'} | \mathbf{x}_{j'-1}, \phi) \right]
\end{aligned}$$

Bisaillon et al., Nonlinear Dynamics, 2015

$$\mathbf{u}_{k+1} = \mathfrak{g}_k \left(\mathbf{u}_k, \phi, \mathbf{f}_k, \mathbf{q}_k^u \right), \tag{7}$$

$$\mathbf{d}_j = \mathfrak{h}_j \left(\mathbf{u}_{d(j)}, \phi, \epsilon_j \right). \tag{8}$$

$$\mathbf{u}_{k+1} = \mathfrak{g}_k \left(\mathbf{u}_k, \phi_k, \mathbf{f}_k, \mathbf{q}_k^u \right), \tag{9}$$

$$\phi_{k+1} = \phi_k + \mathbf{q}_k^{\phi}, \tag{10}$$

$$\mathbf{d}_j = \mathfrak{h}_j \left(\mathbf{u}_{k(j)}, \phi_{k(j)}, \epsilon_j \right) \tag{11}$$

$$\mathbf{x}_k = \begin{Bmatrix} \mathbf{u}_k \\ \phi_k \end{Bmatrix}; \quad \mathbf{q}_k = \begin{Bmatrix} \mathbf{q}_k^u \\ \mathbf{q}_k^{\phi} \end{Bmatrix} \tag{12}$$

Khalil et. al, JSV 2015