Approximate Bayesian Approach to Rapid Structural Identification

A. Lund[(1)], I. Bilionis[(2)], S.J. Dyke[(3)]

[(1)] *Doctoral Candidate in Civil Engineering, Purdue University, alund15@purdue.edu*
[(2)] *Assistant Professor of Mechanical Engineering, Purdue University, ibilion@purdue.edu*
[(3)] *Professor of Mechanical and Civil Engineering, Purdue University, sdyke@purdue.edu*

Abstract

The ability to rapidly assess the condition of a structure in a manner which enables the accurate prediction of its remaining capacity has long been viewed as an essential goal in infrastructure management. Current practice emphasizes visual inspection, in which trained professionals perform routine surveys on structures to estimate their remaining capacity, and threshold-based monitoring, in which sensors are placed in key locations on structures to detect response behavior exceeding predetermined thresholds. Though these methods identify gross structural changes, their ability to rapidly and cost-effectively assess the detailed condition of the structure with respect to its future behavior is limited. Effective strategies for the prediction of future performance center on the identification of dynamic models from observed structural behavior, which is an open research problem.

Bayesian inference techniques give a unique perspective on the identification of nonlinear structural systems as they quantify the inherent epistemic uncertainties that arise due to observations of the system which are both finite in length and limited in the information they contain regarding the states and parameters of interest. However, current applications of Bayesian inference to structural identification are ill-suited to the rapid identification of full-scale structural systems, either due to the simplification of the uncertainties, as with the Kalman filter family of methods, or the computational effort required to form accurate approximations, as with the particle filter family of methods. In this study, we investigate a third family of methods, referred to as variational inference, as an option for nonlinear structural identification. This optimization-based Bayesian inference approach has shown the potential in previous studies to provide detailed estimates of the stochastic models of large-scale nonlinear dynamic systems. We apply variational inference to a simulated nonlinear structural system incorporating Bouc-Wen elements subject to base excitation. Measurement noise and model uncertainties are incorporated in the analysis to assess the robustness of the approach. The results are then compared with those obtained through the unscented Kalman filter to demonstrate the performance of this new method in comparison with commonly used Bayesian identification approaches. The results from this study suggest distinct benefits in terms of the accuracy and robustness of the proposed approach in comparison with the unscented Kalman filter.

Keywords: vibration-based monitoring; Bayesian identification; variational inference; unscented Kalman filter

1. Introduction

Bayesian inference methods have garnered great interest from the structural health monitoring community in recent years for their concise expression of the practical uncertainty inherent to structural systems and their ability to progressively update the dynamic models of these systems from noisy observations of the physical structure. Typically, the inference problem is approached either from an analytical perspective or a sampling perspective. Analytical techniques encompass the Kalman filter [1] and its various approximate forms for nonlinear inference, such as the extended (EKF) [2,3] and unscented (UKF) [4,5] adaptations. These methods have the benefit of computational speed, allowing for near real-time structural identification, but are limited in the assumptions used to generate their analytical framework. Sampling techniques, such as particle filters [6] or sequential Monte Carlo algorithms [7], remove the barriers imposed by this analytical framework, allowing for the accurate inference of complex nonlinear systems. However, this accuracy comes at the cost of increased computational time and limited scalability to larger systems. A more extensive summary of these perspectives and the various techniques which embody them can be found in [8].

Herein we explore variational inference, a third perspective on Bayesian inference which has the potential to strike a balance between the computational speed, accuracy, and scalability of the other techniques. Variational inference uses optimization to approximate the inference of the hidden states, initial conditions, and physical parameters of a system, $\boldsymbol{z} = \boldsymbol{z}_{0:T}$, from observations of its behavior, $\boldsymbol{y} = \boldsymbol{y}_{1:T}$. The algorithm begins with the proposal of a distributional form, $\boldsymbol{Q}$, which serves as an approximation of the posterior density, $p(\boldsymbol{z}|\boldsymbol{y})$. This distributional form, called the *variational family* or *guide*, represents a family of distributions whose members, $q(\boldsymbol{z})$, can be specified by tuning the distributional parameters to different values. Optimization is then performed by finding the member of the variational family which is most similar to the true posterior as defined by the Kullback-Leibler (KL) divergence

$$\begin{aligned}
\hat{q} &= \underset{q \, \in \, \boldsymbol{Q}}{\arg\min} \, \mathrm{KL}\big(q(\boldsymbol{z}) \,||\, p(\boldsymbol{z}|\boldsymbol{y})\big) \\
&= \underset{q \, \in \, \boldsymbol{Q}}{\arg\min} \, \big(\mathbb{E}_{q(\boldsymbol{z})}[\log q(\boldsymbol{z})] - \mathbb{E}_{q(\boldsymbol{z})}[\log p(\boldsymbol{y}, \boldsymbol{z})] + \log p(\boldsymbol{y})\big),
\end{aligned} \quad (1)$$

a measure of the information lost by approximating $p(\boldsymbol{z}|\boldsymbol{y})$ with $q(\boldsymbol{z})$ [9]. Proposing a simple and flexible guide allows for efficient computation of a posterior approximation which is well representative of the true posterior. Variational inference therefore provides a counterpoint to other inference strategies; it is more computationally efficient than the sampling approaches and more flexible in its approximation of the true posterior than the analytical approaches [10].

Variational inference emerged as an alternative Bayesian inference strategy in the mid-90s as a result of the adaptation of mean-field theory from statistical physics [11–14], with the first comprehensive statement of the method introduced in [15]. The generalization of this method to arbitrary guides and large-scale data sets was initially hindered by the need to analytically develop optimization algorithms that depended explicitly on the selected guide [10]. As such, many early works focused on manually expanding the method to common probabilistic models [16–18]. The development of stochastic variational inference [19], which introduced stochastic optimization on mini-batches of data, and black-box variational inference [20], which introduced gradient optimization on Monte Carlo (MC) samples of previously intractable expectations, relieved these issues, allowing for scalable and flexible inference, respectively. Automatic differentiation variational inference (ADVI) represents the current state-of-the-art in variational inference, combining the stochastic and black-box techniques with distributional transformations to enable the automatic implementation of the method on a wide range of probabilistic models [21]. Variational inference has been used extensively in the fields of linguistics [22,23], image processing [24], and computational biology [25,26], among others, but has yet to be adapted for applications in structural health monitoring.

In this paper, we apply ADVI to structural health monitoring through the identification of a dynamical model for a simulated, single degree-of-freedom Bouc-Wen system subject to base vibration. We compare the performance of this inference approach to the unscented Kalman filter (UKF) in terms of the accuracy of

identification, repeatability of the results given variations on the prior information, and robustness to measurement noise and process noise. By so doing, we demonstrate the performance and flexibility of ADVI in comparison with the UKF, and showcase its potential for applications in the modeling and monitoring of structural systems.

2. Case Study – Single Degree-of-Freedom Bouc-Wen System

We are interested in inferring the states, $\boldsymbol{\lambda} = [\boldsymbol{x}, \dot{\boldsymbol{x}}, \boldsymbol{r}]$, and the parameters, $\boldsymbol{\theta} = [c, k, \beta, n, \gamma]$, for the Bouc-Wen system

$$m\ddot{x}(t) + c\dot{x}(t) + kr(t) = -m\ddot{x}_g(t), \tag{2}$$

where the states $\boldsymbol{x}$ and $\dot{\boldsymbol{x}}$ are descriptive of the physical displacement and velocity of the system, respectively, while the state $\boldsymbol{r}$ describes the non-physical Bouc-Wen hysteretic component, the dynamics of which are expressed by

$$\dot{r}(t) = \dot{x}(t) - \beta|\dot{x}(t)||r(t)|^{n-1}r(t) - \gamma\dot{x}(t)|r(t)|^n. \tag{3}$$

The system is driven by a band-limited white noise (BLWN) base excitation sampled at $f_s = 128$ Hz, as shown in Fig. 1(a). The selection of the BLWN excitation is made based on the indications of parameter identifiability from a Sobol' sensitivity analysis [27]. The Sobol' analysis shows that the response of the system to this signal is sensitive to variations in all parameters, though it is significantly more sensitive to the parameters c and k. Further indications of system identifiability are given in Fig. 1(b). This figure shows the response of the system given the parameters $m = 1$ kg, $c = 0.3$ Ns/m, $k = 9$ N/m, $\beta = 2$ m^{-2}, $n = 2$, $\gamma = 1$ m^{-2}, which are selected as the 'true' parameter values for this case study. The response shows that at this input level the system is excited into its nonlinear range of response, and should therefore contain information concerning all parameters of interest.
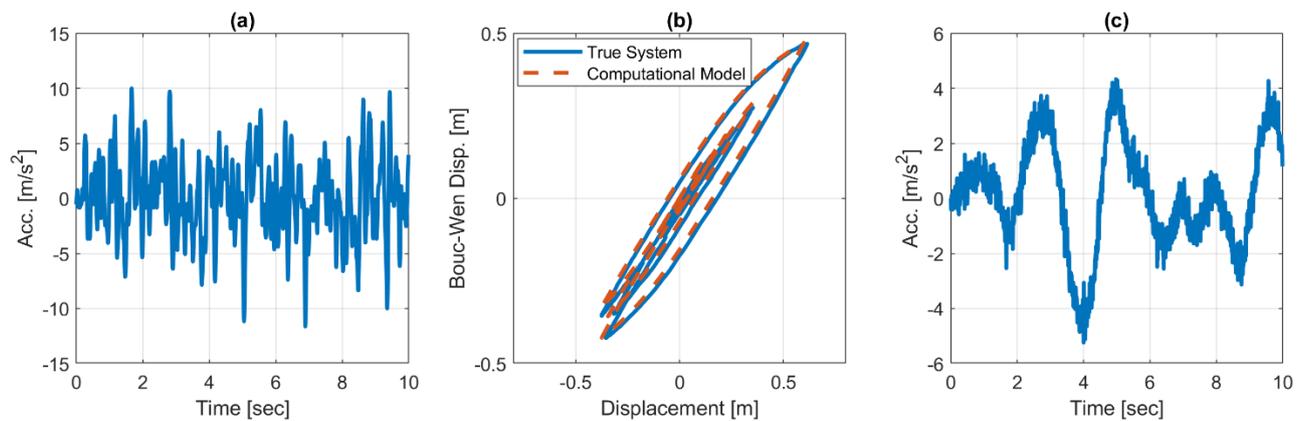

Fig. 1 – Base acceleration and structural responses used in this case study. (a) Base acceleration (b) Bouc-Wen hysteresis (c) Noise-contaminated response acceleration

Nondimensionalizing the equations of motion can ease the computation of the approximate posterior for many inference approaches, as it often puts the states and parameters on a similar scale and allows the inference algorithm to operate in a reduced space of potential solutions [27]. We therefore nondimensionalize Eqs. (2-3) using the time scale $t_c = \sqrt{k/m} = \omega_n$ and the length scale $x_c = 0.05$ m, yielding

$$\ddot{\bar{x}}(\tau) + 2\xi\dot{\bar{x}}(\tau) + \bar{r}(\tau) = -\frac{\ddot{x}_g(\tau/\omega_n)}{x_c\omega_n^2} \tag{4}$$

$$\dot{\bar{r}}(\tau) = \dot{\bar{x}}(\tau) - \beta x_c^n|\dot{\bar{x}}(\tau)||\bar{r}(\tau)|^{n-1}\bar{r}(\tau) - \gamma x_c^n\dot{\bar{x}}(\tau)|\bar{r}(\tau)|^n. \tag{5}$$

3

The problem is now transformed such that the states and parameters we wish to infer are $\bar{\boldsymbol{\lambda}} = \left[\overline{x}, \dot{\overline{x}}, \overline{r}\right] = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3]$ and $\overline{\boldsymbol{\theta}} = [\xi, \omega_n, \beta, n, \gamma] = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$. We will refer to these collectively as the hidden states of the system, $\boldsymbol{z} = \left[\bar{\boldsymbol{\lambda}}, \overline{\boldsymbol{\theta}}\right]$.

Inference methods are typically structured to work with systems having discrete relationships among the states and observations. Though some methods have been designed to take advantage of continuous dynamical models, we have elected to use an Euler discretization scheme on Eqs. (4-5) to preserve the uniformity of the comparison between the UKF and variational inference methods. Likewise, in both cases we choose to observe the system acceleration, shown in Fig. 1(c),

$$y(\tau) = -2\xi\dot{\overline{x}}(\tau) - \overline{r}(\tau). \tag{6}$$

In this case study, we explore the accuracy of the inference methods given model uncertainty and measurement error by adding zero-mean Gaussian distributed process noise and measurement noise to the state transition and observation equations. For a base exploration of the relative capabilities of the two methods, we impose process noises with standard deviations representing a 1%, 2%, and 2% root-mean-square signal-to-noise ratio (RMS-SNR) on the displacement, velocity, and Bouc-Wen displacement, respectively. Note that this noise is imposed on the states according to

$$w_i \sim N\left(0, \left(\sqrt{\Delta\tau} \cdot w_{\text{RMS-SNR}} \cdot \text{RMS}(\boldsymbol{x}_i)\right)^2\right). \tag{7}$$

The difference between the true response, which we refer to as the one containing process noise, and the model response can be seen in Fig. 1(b). We also impose a measurement noise with a standard deviation representing a 20% RMS-SNR on the acceleration, which can be described by the distribution

$$v \sim N\left(0, \left(v_{\text{RMS-SNR}} \cdot \text{RMS}(y)\right)^2\right). \tag{8}$$

Small variations of the prior distributions used to inform these inference methods can lead to large changes in the identified states. We therefore evaluate the robustness of the UKF and variational inference methods by proposing 50 distributions on the prior assumptions for the parameters which are representative of likely assumptions that an experimentalist might make in defining the problem. This evaluation is done by first selecting a Latin Hypercube Sample (LHS) of the means of the prior distributions. For both algorithms we represent the parameters ω_n, β, and γ as lognormally distributed and use LHS to select their means from the ranges $\ln(\omega_n) \epsilon [0,2.3]$, $\ln(\beta) \in [0,3.2]$, and $\ln(\gamma) \in [0,3.2]$. As the UKF assumes all states are Gaussian distributed, we log-transform the parameters themselves and perform the inverse transform on the inferred values to discern the identified parameters. Variational inference is more flexible and the lognormal distributions on these parameters can be used directly. The priors on the parameters ξ and n are treated differently between the UKF and variational inference algorithms to take advantage of the additional flexibility of variational inference. For the UKF these parameter priors are represented by lognormal distributions with means selected from the ranges $\ln(\xi) \epsilon [-2.3,0]$ and $\ln(n) \epsilon [0.7,1.8]$. For variational inference, we take advantage of the clear domains of these parameters ($\xi \epsilon [0,1]$ and $n \epsilon [1,6]$) and represent their prior distributions as uniform on a subset of that domain. The left edge of the distribution is held constant for each case at 0 for ξ and 2 for n. The right edge is varied according to the same LHS as was used to define the mean of the UKF implementation of these parameters, yielding a range of $\xi \epsilon [0.1,1]$ and $n \epsilon [2,6]$. After establishing the means of these distributions through LHS, the variance for each prior is hand-selected to represent the uncertainty level a typical experimentalist might assign to the parameter before identification of the system. Prior distributions on the dynamic states are uniformly set as $x_i(0) \sim N(0, 0.25^2)$ for each inference trial on the parameter priors, as the system is known to be at rest prior to excitation. The identification trials resulting from these prior distributions will be analyzed in detail in Section 4.

3. Methodology and Implementation

The objective of Bayesian inference for system identification is to discern the hidden states, $\mathbf{z} = \mathbf{z}_{0:T}$, of a dynamical system from noisy observations of its behavior, $\mathbf{y} = \mathbf{y}_{1:T}$, where the form of the system model is known to be a sufficient, but imperfect, representation of the physical system. Systems of this nature are typically described by a Markovian transmission probability $p(\mathbf{z}_t|\mathbf{z}_{t-1})$ and an emission probability $p(\mathbf{y}_t|\mathbf{z}_t)$. We can therefore discern the posterior probability $p(\mathbf{z}|\mathbf{y})$ by using either a batch method, which infers the posterior from all data points simultaneously

$$p(\mathbf{z}_{0:T}|\mathbf{y}_{1:T}) = \frac{\prod_{t=0}^{T-1} p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}) \prod_{t=0}^{T-1} p(\mathbf{z}_{t+1}|\mathbf{z}_t) \, p(\mathbf{z}_0)}{p(\mathbf{y}_{1:T})}, \tag{9}$$

or a filtering method, which recursively constructs the posterior by forming a prediction of the current state with the Chapman-Kolmogorov equation,

$$p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t}) = \int p(\mathbf{z}_{t+1}|\mathbf{z}_t) p(\mathbf{z}_t|\mathbf{y}_{1:t}) \, d\mathbf{z}_t, \tag{10}$$

and correcting that prediction with Bayes' Theorem,

$$p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t+1}) = \frac{p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}) p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}. \tag{11}$$

The UKF is a well-recognized filtering method which has been used to successfully identify the dynamic states and parameters of a variety of nonlinear structural systems, including the Bouc-Wen system which is the focus of this study [28,29]. Variational inference, in contrast, traditionally operates in batch form. The details of these inference approaches are discussed in the following sections.

3.1 Unscented Kalman Filter

Kalman filters provide an analytical solution to Eqs. (10-11) given that the system is linear and that the parameters describing the model uncertainty and measurement uncertainty are normally distributed [1]. The UKF expands the Kalman filter to deal with nonlinear systems by forming an approximation of the posterior from the statistics of a set of deterministically selected sigma points, $\boldsymbol{\mathcal{X}}^{(t)}$, which are passed through the nonlinear system equations [5].

The filter first requires the definition of the prior distribution on the hidden states, $\mathbf{z}_0 \sim N(\boldsymbol{\mu}_0, \mathbf{P}_0)$, and the nonlinear transmission and emission probabilities on those states, which for our case can be expressed as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \overline{\boldsymbol{\theta}} \end{bmatrix}_{\tau+1} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \overline{\boldsymbol{\theta}} \end{bmatrix}_{\tau} + \Delta\tau \begin{bmatrix} x_2 \\ -2\theta_1 x_2 - x_3 - \dfrac{\ddot{x}_g(\tau/\theta_2)}{x_c \theta_2} \\ x_2 - \theta_3 x_c^{\theta_4} |x_2||x_3|^{\theta_4-1} x_3 - \gamma x_c^{\theta_4} x_2 |x_3|^{\theta_4} \\ \mathbf{0} \end{bmatrix}_{\tau} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \mathbf{0} \end{bmatrix}_{\tau} \tag{12}$$

$$\mathbf{y}_{\tau+1} = -2\theta_{1,\tau+1} x_{2,\tau+1} - x_{3,\tau+1} + v_{\tau+1}. \tag{13}$$

Each filter iteration then begins with the determination of a weighted set of sigma points which are selected to describe the distribution of the states, $\mathbf{z}_\tau$. The selection of sigma points can be accomplished in a number of ways to emphasize different aspects of the generating distribution. The sigma point set used herein is constructed using an augmented system state, $\mathbf{z}_\tau^a = [\mathbf{z}_\tau \quad \mathbf{w}_\tau \quad \mathbf{v}_\tau]^T$, such that $\boldsymbol{\mathcal{X}}^{(\tau)} = [\boldsymbol{\mathcal{X}}^{(\mathbf{z}_\tau)} \quad \boldsymbol{\mathcal{X}}^{(\mathbf{w}_\tau)} \quad \boldsymbol{\mathcal{X}}^{(\mathbf{v}_\tau)}]^T$, as described in [28]. These sigma points are then propagated through the state transition model given by Eq. (12) to estimate $p(\mathbf{z}_{\tau+1}|\mathbf{y}_{1:\tau})$, which is the predicted distribution of the states given prior information about the system and its measurement. Next, the predicted state distribution is projected through the observation function by Eq. (13) to estimate the distribution on the predicted

5

measurement. This distribution is then compared with the true measurement to generate the intermediate posterior distribution on the hidden states, $p(\mathbf{z}_{\tau+1}|\mathbf{y}_{1:\tau+1})$. For a more comprehensive explanation of the UKF algorithm, the reader is referred to [30].

3.2 Automatic Differentiation Variational Inference

ADVI automates variational inference by transforming the problem into a domain compatible with a generalized solution approach. The procedure for this method can be understood in terms of the four steps given in Fig. 2.

Fig. 2 – Key steps of the ADVI algorithm

In the first step, defining problem domain, we make explicit our understanding of the stochastic dynamical system by defining the prior distributions on the hidden states, the transmission probabilities, and the emission probabilities. For this case study, the prior distributions on the dynamical states are specified as $x_i(0) \sim N(0, 0.25^2)$ and the prior distributions on the parameters are given by $\xi \sim Uniform(0, \mu_\xi)$, $\omega_n \sim Lognormal(\mu_{\omega_n}, \sigma_{\omega_n}^2)$, $\beta \sim Lognormal(\mu_\beta, \sigma_\beta^2)$, $n \sim Uniform(0, \mu_n)$, and $\gamma \sim Lognormal(\mu_\gamma, \sigma_\gamma^2)$. The transition probabilities among the states can then be defined as

$$p(x_1(\tau+1)|\bar{\lambda}(\tau), \overline{\boldsymbol{\theta}}) = N(x_1(\tau+1) \mid x_1(\tau) + \Delta\tau \cdot x_2(\tau), w_1^2)$$

$$p(x_2(\tau+1)|\bar{\lambda}(\tau), \overline{\boldsymbol{\theta}}) = N\left(x_2(\tau+1) \mid x_2(\tau) + \Delta\tau\left(-2\theta_1 x_2(\tau) - x_3(\tau) - \frac{\ddot{x}_g(\tau/\theta_2)}{x_c\theta_2}\right), w_2^2\right)$$

$$p(x_3(\tau+1)|\bar{\lambda}(\tau), \overline{\boldsymbol{\theta}}) =$$

$$N\left(x_3(\tau+1) \mid x_3(\tau) + \Delta\tau\left(x_2(\tau) - \theta_3 x_c^{\theta_4}|x_2(\tau)||x_3(\tau)|^{\theta_4-1}x_3(\tau) - \gamma x_c^{\theta_4}x_2(\tau)|x_3(\tau)|^{\theta_4}\right), w_3^2\right), \quad (14)$$

and the emission probability between the states and the observations can be defined as

$$p(y(\tau+1)|\bar{\lambda}(\tau+1), \overline{\boldsymbol{\theta}}) = N(y(\tau+1) \mid -2\theta_1 x_2(\tau+1) - x_3(\tau+1), v^2). \quad (15)$$

Given these probabilities, we can express the joint probability of the hidden states and observations as

$$p(\mathbf{y}, \bar{\lambda}, \overline{\boldsymbol{\theta}}) = \prod_{\tau=0}^{T-1} p(\mathbf{y}_{\tau+1}|\bar{\lambda}_{\tau+1}, \overline{\boldsymbol{\theta}}) \prod_{\tau=0}^{T-1} p(\bar{\lambda}_{\tau+1}|\bar{\lambda}_\tau, \overline{\boldsymbol{\theta}}) \prod_{i=1}^{3} p(x_i(0)) \prod_{i=1}^{5} p(\theta_i). \quad (16)$$

To automate the solution approach, the prior distributions on the hidden parameters $\overline{\boldsymbol{\theta}}$ are transformed to have support on the Euclidean space $\mathbb{R}^K$. The joint density is then expressed in terms of transformed parameters $\boldsymbol{\zeta} = T(\overline{\boldsymbol{\theta}})$, such that

$$p(\mathbf{y}, \bar{\lambda}, \boldsymbol{\zeta}) = p\left(\mathbf{y}, \bar{\lambda}, T^{-1}(\boldsymbol{\zeta})\right) |\det J_{T^{-1}}(\boldsymbol{\zeta})|, \quad (17)$$

where $T(\cdot)$ is a one-to-one differentiable function which transforms $\overline{\boldsymbol{\theta}}$ to have full support in $\mathbb{R}^K$.

In the second step, proposing a variational family, we specify the form of the approximate posterior model that we believe will adequately represent the system. As we have transformed the hidden states and parameters to have support in $\mathbb{R}^K$, the Gaussian distribution is a simple choice for the variational family. We represent the approximate posterior of the displacement, velocity, and Bouc-Wen displacement of the system to be mutually independent and capture the Markovian transitions within each state through a tri-diagonal covariance matrix. This yields the variational families of

$$q(\bar{\lambda}; \boldsymbol{\phi}_{\bar{\lambda}}) = N(\bar{\lambda}|\boldsymbol{\mu}_{\bar{\lambda}}, \mathbf{L}_{\bar{\lambda}}\mathbf{L}_{\bar{\lambda}}^T) \quad (18)$$

6

which are parameterized by the variational parameters $\boldsymbol{\phi}_{\bar{\lambda}} = (\boldsymbol{\mu}_{\bar{\lambda}}, \mathbf{L}_{\bar{\lambda}})$, where $\mathbf{L}_{\bar{\lambda}}$ is a lower triangular matrix with non-zero elements only on the two primary diagonals. These variational parameters are unconstrained in $\mathbb{R}^{3(3\mathrm{T}-1)}$. Likewise, we propose a Gaussian variational family on the hidden parameters which expresses their mutual independence through a diagonal covariance matrix, yielding

$$q(\boldsymbol{\zeta}; \boldsymbol{\phi}_{\boldsymbol{\zeta}}) = N\left(\boldsymbol{\zeta} | \boldsymbol{\mu}_{\boldsymbol{\zeta}}, \mathrm{diag}\left(\exp(\boldsymbol{\rho}_{\boldsymbol{\zeta}})^2\right)\right) \tag{19}$$

which is parameterized by $\boldsymbol{\phi}_{\boldsymbol{\zeta}} = (\boldsymbol{\mu}_{\zeta,1}, \dots, \boldsymbol{\mu}_{\zeta,K}, \boldsymbol{\rho}_{\zeta,1}, \dots, \boldsymbol{\rho}_{\zeta,K})$ and produces a set of variational parameters which are unconstrained in $\mathbb{R}^{2K}$. The full model then requires optimization on $\mathbb{R}^{3(3\mathrm{T}-1)+2K}$.

In the third step, specifying the loss function, we set up our optimization problem. Ideally, we would minimize the difference between the true and approximate posteriors using the KL divergence given in Eq. (1). However, the KL divergence explicitly depends on the model evidence $p(\mathbf{y}_{1:T})$ which we do not know. Instead, we maximize the evidence lower bound (ELBO)

$$\begin{aligned}\mathrm{ELBO}(q) &= \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{y}|\mathbf{z})] + \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{z})] - \mathbb{E}_{q(\mathbf{z})}[\log q(\mathbf{z})] \\ &= \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{y}|\mathbf{z})] - \mathrm{KL}(q(\mathbf{z}) \,||\, p(\mathbf{z})),\end{aligned} \tag{20}$$

which is equivalent to the negative of the KL divergence, Eq. (1), plus $\log p(\mathbf{y})$. The ELBO can be understood from its two components to strike a balance between encouraging densities which fit to the observed data and encouraging densities which stay close to the prior. When this balance is achieved, the optimal density will express the behavior of the true system without overfitting to the limited data set used for optimization. For our problem, the ELBO can be expressed as

$$\mathrm{ELBO}(q) = \mathbb{E}_{q(\bar{\lambda},\zeta;\phi)}\left[\log p\left(\mathbf{y}, \bar{\lambda}, T^{-1}(\zeta)\right) + \log|\det J_{T^{-1}}(\theta)|\right] - \mathbb{E}_{q(\bar{\lambda},\zeta;\phi)}[\log q(\bar{\lambda}, \zeta; \phi)]. \tag{21}$$

In the final step, performing stochastic optimization, our goal is to use a noisy estimate of the gradient of the ELBO to walk toward locally optimal values of the hidden states. However, it is difficult to take the gradient of the ELBO directly, as the ELBO involves an intractable expectation. To resolve this issue and allow for the use of automatic differentiation to evaluate the gradient, we introduce an additional transformation on the parameters, referred to as elliptical standardization. This transformation can be expressed as $\eta_{\bar{\lambda}} = S_{\phi_{\bar{\lambda}}}(\bar{\lambda}) = \mathbf{L}_{\bar{\lambda}}^{-1}(\bar{\lambda} - \boldsymbol{\mu}_{\bar{\lambda}})$ for the states and $\eta_{\zeta} = S_{\phi_{\zeta}}(\zeta) = \mathrm{diag}(\exp(\boldsymbol{\rho}_{\zeta}))^{-1}(\zeta - \boldsymbol{\mu}_{\zeta})$ for the parameters, yielding a modified ELBO

$$\begin{aligned}\mathrm{ELBO}(q) = \mathbb{E}_{N(\eta;0,I)}\Bigg[&\log p\left(\mathbf{y}, S_{\phi_{\bar{\lambda}}}^{-1}(\eta_{\bar{\lambda}}), T^{-1}\left(S_{\phi_{\zeta}}^{-1}(\eta_{\zeta})\right)\right) + \log\left|\det J_{T^{-1}}\left(S_{\phi_{\zeta}}^{-1}(\eta_{\zeta})\right)\right|\Bigg] \\ &- \mathbb{E}_{q(\bar{\lambda},\zeta;\phi)}[\log q(\bar{\lambda}, \zeta; \phi)],\end{aligned} \tag{22}$$

which allows us to use Monte Carlo methods (typically with only 1 sample) to obtain a noisy approximation of the ELBO for the automatic evaluation of the gradient. Stochastic optimization can then be performed using a number of algorithms. In this study, we implement ADVI using the ADAM stochastic optimization algorithm as part of the python library pyTorch [31]. For further information about variational inference and ADVI, the reader is referred to [10,21].

4. Results and Discussion

As discussed in Section 2, our goal in this case study is to infer the hidden states $\mathbf{z} = [\bar{\lambda}, \bar{\theta}]$ of the nondimensionalized Bouc-Wen system from noisy observations of its response to a BLWN base excitation. 50 identification trials are conducted with the UKF and variational inference methods in order to evaluate both the accuracy and the robustness of the identification. Final identified parameters from each approach are used to re-simulate the response of the Bouc-Wen system. The identification trials which result in the minimum

7

RMS error on the states are shown in Fig. 3 in terms of the prior and posterior distributions on the parameters. The results show that the variational inference algorithm (Fig. 3(f-j)) yields a more confident and accurate posterior distributions on the parameters, particularly for the nonlinear parameters of $\beta, n,$ and γ.

The increased accuracy of the posterior means for the variational inference case do not translate to a significantly improved estimate of the states, as demonstrated in Fig. 4. This figure shows the true response of the Bouc-Wen system in comparison with model responses which are re-simulated from the posterior modes given for the UKF and variational inference methods in Fig. 3. The results confirm those of the Sobol sensitivity analysis at the beginning of this study, which indicated that the response of the system is much less sensitive to variations in the nonlinear parameters than variations in the linear parameters. It appears that if the linear parameters are identified with reasonable accuracy, several locally optimal solutions for the nonlinear parameters exist which can enhance the similarity of the identified system to the true response.
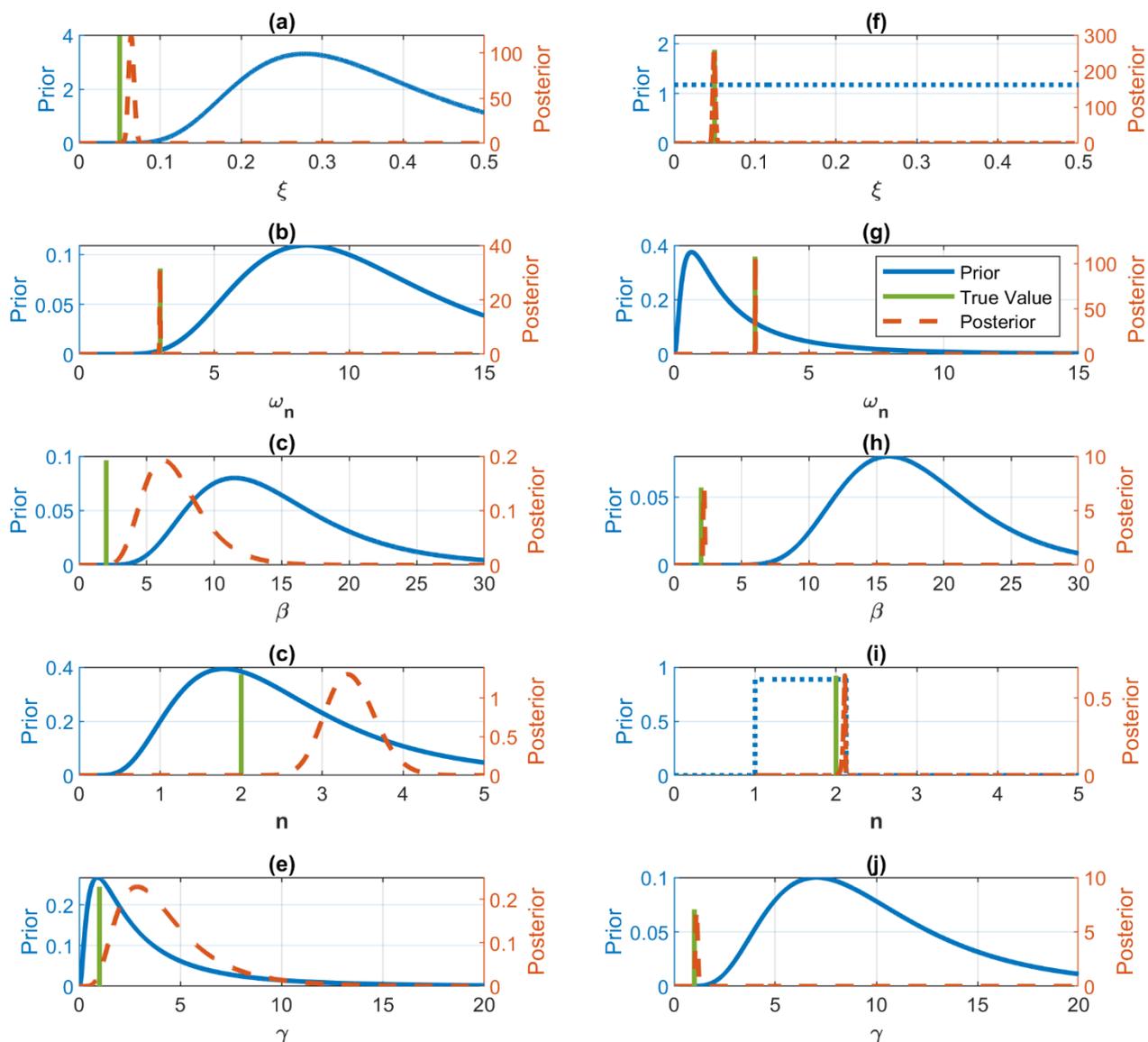

Fig. 3 - Priors and posteriors of the minimum RMS error identification trials. (a-e) UKF identification trials (f-j) variational inference identification trials
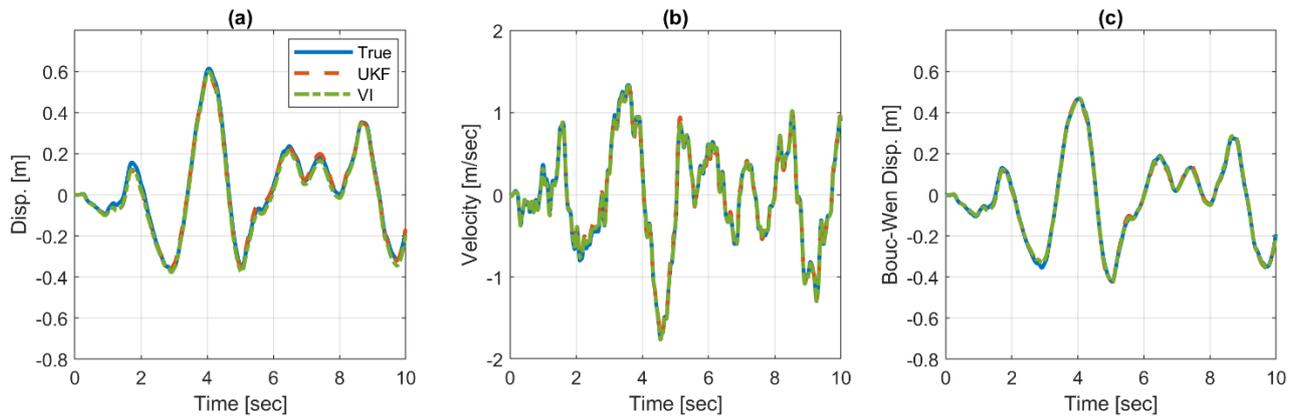
8

Fig. 4 - Comparison of the true system response with model responses re-simulated from the minimum RMS error identification trials

We further the comparison between the UKF and variational inference methods by analyzing their relative accuracy over all 50 identification trials, as shown in Fig. 5 and Fig. 6. Fig. 5 gives the distribution of posterior means for the identification trials. The most prominent feature of these data is the change in the spread of posterior means from the linear parameters ξ and ω_n to the nonlinear parameters β, n, and γ, regardless of the inference approach used to identify them. This result once again confirms limited sensitivity of the system response to the nonlinear parameters relative to the linear parameters. Comparing the two inferences methods, we see that for all parameters the variational inference method exhibits a much smaller spread in identified values, which is focused on the true value of the parameters. Looking at the nonlinear parameters specifically, we see that even though there is a greater spread of the variational inference posteriors they are still fairly well concentrated when compared with the UKF posteriors, which exhibit significant concentrations of outliers.

To understand what these variations in the posterior mean for the quality of the inferred model, we turn to Fig. 6, which provides a case-by-case comparison of the RMS error in the states which have been re-simulated with parameters set to the posterior modes of either the UKF or variational inference identification trial. The data clearly show that regardless of variations in the parameters, the variational inference method consistently provides a low-error response with respect to the true states. The UKF is able to match this performance for the majority of the identification trials, but experiences larger variations in error due to the outlying parameter cases shown in Fig. 5. These results suggest that even though the posteriors may not precisely match the true parameters, the variational inference approach is more adept at consistently finding parameter combinations which locally minimize the error between the true response and the re-simulated model.

5. Conclusions

In this case study we compare the ability of the UKF and variational inference methods to identify the hidden states and parameters of a simulated single degree-of-freedom Bouc-Wen system excited by a BLWN base motion. 50 identification trials were performed to analyze the relative accuracy and robustness of the two methods. The results indicate that variational inference consistently provides low-error models of the system in comparison with the UKF method. However, the robustness of this method does come at the cost of some increased computational time. Whereas the UKF executes on the order of seconds, variational inference requires execution times on the order of hours. As this issue is addressed in future implementations of variational inference, it will become an even more valuable option for structural health monitoring applications.
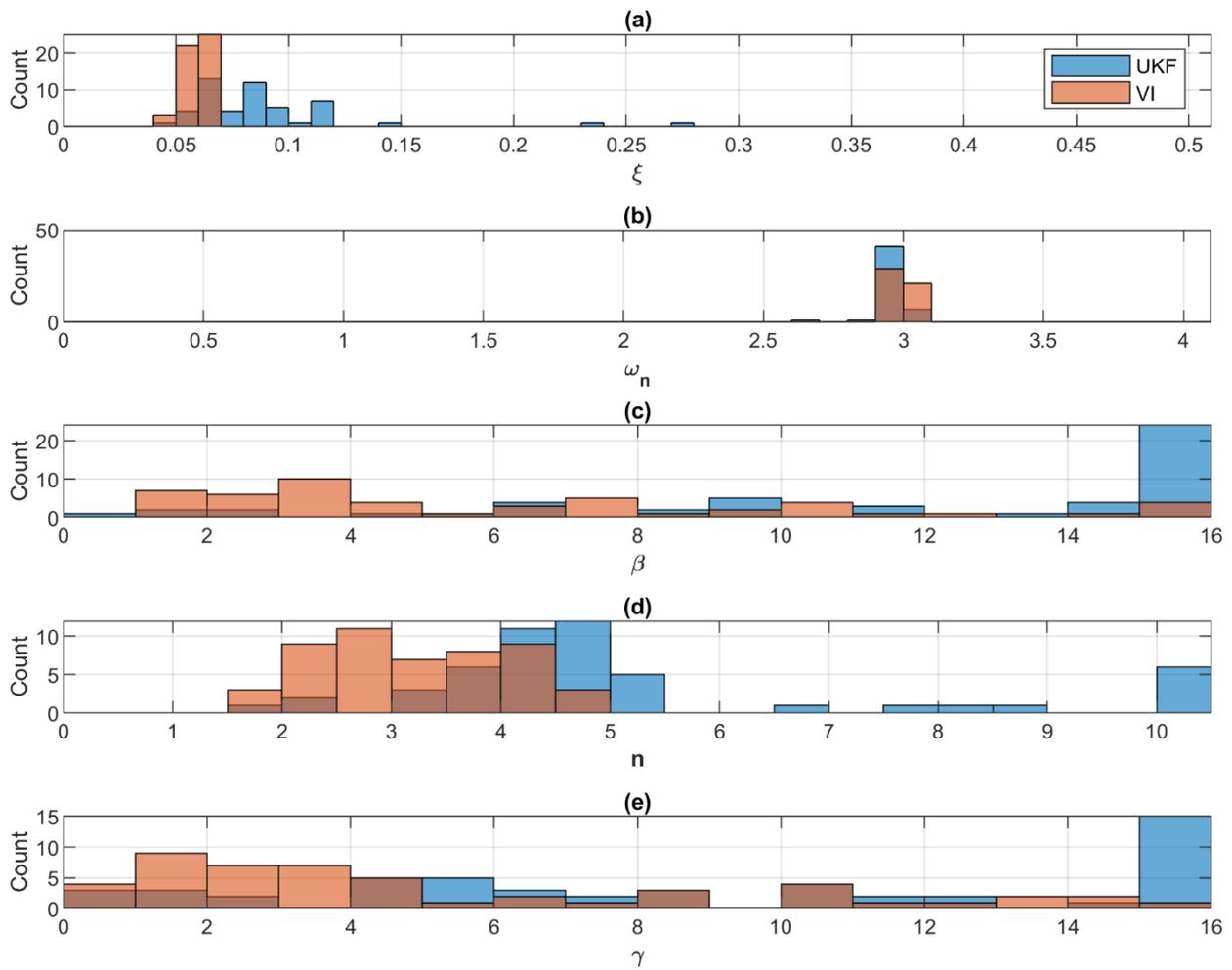
Fig. 5 - Distribution of the posterior means on the parameters for all 50 identification trials
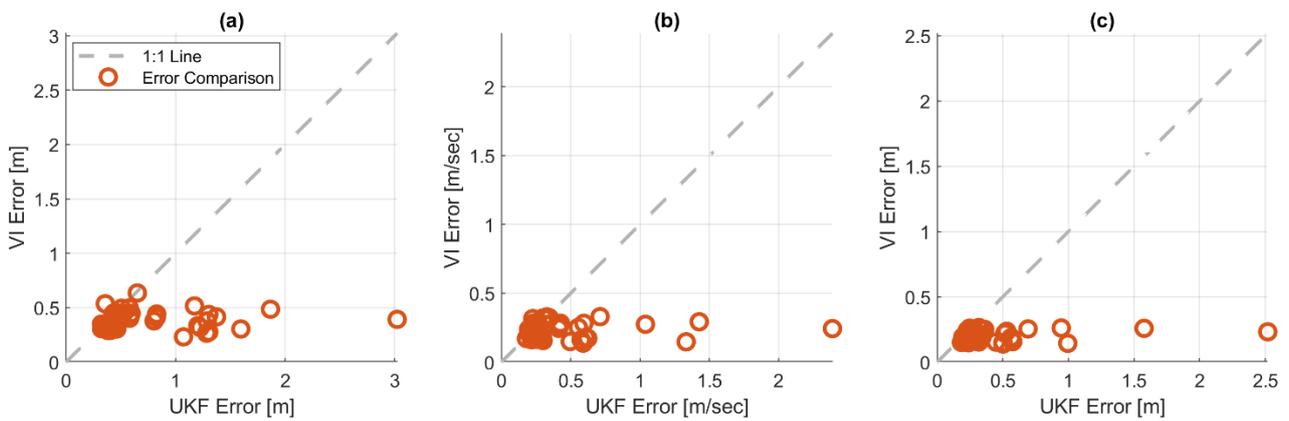

Fig. 6 - Comparison of the RMS error on the states for the 50 identification trials

10

6. Acknowledgements

7. References

[1]   Kalman RE (1960): A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82** (1) 35–45. https://doi.org/10.1115/1.3662552.

[2]   Schmidt SF (1981): The Kalman filter - Its recognition and development for aerospace applications. *Journal of Guidance, Control, and Dynamics*, **4** (1) 4–7. https://doi.org/10.2514/3.19713.

[3]   Sorenson HW (1966): Kalman filtering techniques. *Advances in Control Systems*, **3** 219–292. https://doi.org/10.1016/b978-1-4831-6716-9.50010-2.

[4]   Julier SJ, Uhlmann JK, Durrant-Whyte HF (1995): A new approach for filtering nonlinear systems. *Proceedings of American Control Conference*, Seattle, Washington. https://doi.org/10.1109/ACC.1995.529783.

[5]   Wan EA, Van Der Merwe R (2000): The unscented Kalman filter for nonlinear estimation. *IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, Lake Louise, Alberta, Canada.

[6]   Gordon NJ, Salmond DJ, Smith AFM (1993): Novel approach to nonlinear / non-Gaussian Bayesian state estimation. *IEE Proceedings (Radar and Signal Processing)*, **140** (2) 107–113.

[7]   Doucet A, Godsill S, Andrieu C (2000): On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10** (3) 197–208.

[8]   Haug AJ (2012): Bayesian Estimation and Tracking: A Practical Guide, John Wiley & Sons, 1st ed.

[9]   Bishop CM (2006): Pattern recognition and machine learning, Springer, 1st ed. https://doi.org/10.1117/1.2819119.

[10]   Blei DM, Kucukelbir A, Mcauliffe JD (2017): Variational inference: A review for statisticians. *Journal of the American Statistical Association*, **112** (518) 859–877. https://doi.org/10.1080/01621459.2017.1285773.

[11]   Saul LK, Jaakkola T, Jordan MI (1996): Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research*, **4** 61–76.

[12]   Saul LK, Jordan MI (1996): Exploiting Tractable Substructures in Intractable Networks. *Advances in Neural Information Processing Systems*.

[13]   Jaakkola TS, Jordan MI (1996): A variational approach to Bayesian logistic regression models and their extensions. *6th International Workshop on Artificial Intelligence and Statistics*.

[14]   Hinton GE, Van Camp D (1993): Keeping Neural Networks Simple by Minimizing the Description Length of the Weights. *6th Annual Conference on Computational Learning Theory*.

[15]   Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK (1999): An Introduction to Variational Methods for Graphical Models. *Machine Learning*, **37** 183–233. https://doi.org/10.1007/978-94-011-5014-9_5.

[16]   Blei DM, Ng AY, Jordan MI (2003): Latent Dirichlet allocation. *Journal of Machine Learning Research*, **3** 993–1022.

[17]   Barber D, Wiegernick W (1999): Tractable Variational Structures for Approximating Graphical Models. *Advances in Neural Information Processing Systems*.

[18]   Beal MJ 2003, Variational Algorithms for Approximate Bayesian Inference, University College, London.

[19]   Hoffman MD, Blei DM, Wang C, Paisley J (2013): Stochastic variational inference. *Journal of Machine Learning Research*, **14** (1) 1303–1347.

[20]   Ranganath R, Gerrish S, Blei DM (2014): Black box variational inference. *Artificial Intelligence and Statistics*,

11

814–822.

[21] Kucukelbir A, Tran D, Ranganath R, Gelman A, Blei DM (2017): Automatic differentiation variational inference. *Journal of Machine Learning Research*, **18** (1) 430–474.

[22] Blei DM, Lafferty JD (2007): A Correlated Topic Model of Science. *The Annals of Applied Statistics*, **1** (1) 17–35. https://doi.org/10.1214/07-AOAS114.

[23] Cohen SB, Smith NA (2010): Covariance in unsupervised learning of probabilistic grammars. *Journal of Machine Learning Research*, **11** 3017–3051.

[24] Likas A, Galatsanos NP (2004): A variational method for Bayesian blind image deconvolution. *International Conference on Image Processing*. https://doi.org/10.1109/ICIP.2003.1246846.

[25] Logsdon BA, Hoffman GE, Mezey JG (2010): A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis. *BMC Bioninformatics*, **11** (1) 58–70.

[26] Raj A, Stephens M, Pritchard JK (2014): fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, **197** (2) 573–589. https://doi.org/10.1534/genetics.114.164350.

[27] Lund A, Dyke SJ, Song W, Bilionis I (2019): Global sensitivity analysis for the design of nonlinear identification experiments. *Nonlinear Dynamics*, **98** (1) 375–394. https://doi.org/10.1007/s11071-019-05199-9.

[28] Wu M, Smyth AW (2007): Application of the unscented Kalman filter for real-time nonlinear structural system identification. *Structural Control and Health Monitoring*, **14** 971–990. https://doi.org/10.1002/stc.

[29] Chatzi EN, Smyth AW, Masri SF (2010): Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty. *Structural Safety*, **32** 326–337. https://doi.org/10.1016/j.strusafe.2010.03.008.

[30] Sarkka S (2013): Bayesian filtering and smoothing, Cambridge University Press, 1st ed. https://doi.org/10.1017/CBO9781139344203.

[31] Kingma DP, Ba J (2014): Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.